

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2004, 10:55:53 / Search time 3023 Seconds
(without alignments)
9850.038 Million cell updates/sec

Title: US-09-976-673-11

Perfect score: 687

Sequence: 1 atggtgagcgcgcctgtgaa.....tgcgcgagaagcccaactga 687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pac.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_stb.*
- 12: gb_ey.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pac.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_hcg_hum.*
- 31: em_hcg_inv.*
- 32: em_hcg_other.*
- 33: em_hcg_mus.*
- 34: em_hcg_pln.*
- 35: em_hcg_rtd.*
- 36: em_hcg_mam.*
- 37: em_hcg_vtc.*
- 38: em_ey.*
- 39: em_hcgo_hum.*
- 40: em_hcgo_mus.*
- 41: em_hcgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	100.0	687	6	AX527898 Sequence
2	671	97.7	687	6	AX527900 Sequence
3	671	97.7	1396	6	AX527902 Sequence
4	669.4	97.4	1424	6	AX527904 Sequence
5	441.2	64.2	681	6	AX527894 Sequence
6	439.6	64.0	681	6	AX527896 Sequence
7	438.2	63.8	684	6	AX527910 Sequence
8	436.6	63.6	684	6	AX527892 Sequence
9	436.6	63.6	684	3	AF353776 Heteracti
10	435	63.3	910	6	AX527888 Sequence
11	435	63.3	910	6	AX527914 Sequence
12	431.8	62.9	908	6	AX527890 Sequence
13	429.2	62.5	680	6	AX527912 Sequence
14	424.8	61.8	699	12	AY233273 Synthetic
15	417.8	60.8	707	6	AX824739 Sequence
16	396.8	57.8	684	6	AX824741 Sequence
17	362.6	52.8	681	6	AX573118 Sequence
18	361.2	52.6	835	3	AF363775 CondyIact
19	361.2	52.6	835	6	AX573116 Sequence
20	358	52.1	684	3	AF383155 CondyIact
21	351.2	51.1	835	3	AY037777 CondyIact
22	306	44.5	678	12	AF506027 Synthetic
23	299	43.5	713	6	AX233627 Sequence
24	298.8	43.5	681	6	AX233584 Sequence
25	298.2	43.4	681	12	AF506025 Synthetic
26	298.2	43.4	1395	12	AF506026 Synthetic
27	295	42.9	10141	12	AY342347 Red H-Pel
28	294	42.8	10276	12	AY342348 Red H-Sci
29	292.4	42.6	678	6	AX370404 Sequence
30	292.4	42.6	678	6	AX370406 Sequence
31	292.4	42.6	678	6	AX824725 Sequence
32	291.4	42.4	4692	6	AX463702 Sequence
33	291.4	42.4	9320	6	AX663075 Sequence
34	290.8	42.3	6893	6	AX823860 Sequence
35	289.2	42.1	678	6	AX370408 Sequence
36	289.2	42.1	678	6	AX824732 Sequence
37	288.4	42.0	678	6	AX824729 Sequence
38	288.2	42.0	1050	6	AX666133 Sequence
39	287.6	41.9	675	6	AX824731 Sequence
40	268	39.0	955	3	AF246709 Anemonia
41	266.4	38.8	699	3	AF322222 Anemonia
42	264.8	38.5	696	6	AX668890 Sequence
43	264.8	38.5	696	6	AX824727 Sequence
44	261.2	38.0	720	6	AX666128 Sequence
45	261.2	38.0	9258	12	AF383623 Cloning v

ALIGNMENTS

RESULT 1
AX527898 687 bp DNA linear PAT 21-NOV-2002
LOCUS Sequence 11 from Patent WO0230965.
DEFINITION AX527898
ACCESSION AX527898
VERSION AX527898.1 GI:25172348
KEYWORDS
SOURCE Heteractis crispa (leathery sea anemone)
ORGANISM Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
Stichodactylidae; Heteractis.

REFERENCE 1
AUTHORS Lukyanov,S.A., Pradkov,A.F., Lukyanov,K.A. and Gurskaya,N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 11 18-APR-2002;

FEATURES
source Clontech Laboratories Inc. (US)
Location/Qualifiers
1..687
/organism="Heteractis criepa"
/mol_type="unassigned DNA"
/db_xref="taxon:15771"

ORIGIN

Query Match 100.0%; Score 687; DB 6; Length 687;
Best Local Similarity 100.0%; Pred. No. 2.2e-79;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGAGCGGCTCTGTAAGAGAGATATGCGCATCAAGTGTACATGAGGGACCGTG 60
Db 1 ATGTGAGCGGCTCTGTAAGAGAGATATGCGCATCAAGTGTACATGAGGGACCGTG 60
QY 61 AACGGCACTACTTCAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
Db 61 AACGGCACTACTTCAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 121 AGCATGAAATCCAGTGAACGAGGGGCGCCCTTGCTTGCTTGCTTGCTTGCTTGCT 180
Db 121 AGCATGAAATCCAGTGAACGAGGGGCGCCCTTGCTTGCTTGCTTGCTTGCTTGCT 180
QY 181 CCTGTGCGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db 181 CCTGTGCGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 241 TTCAAGCAGAGCTTCCCGAGGGGCTTCACTGAGGAGAGAACCAACCACTGAGAGGAG 300
Db 241 TTCAAGCAGAGCTTCCCGAGGGGCTTCACTGAGGAGAGAACCAACCACTGAGAGGAG 300
QY 301 GGCATCTTACCGCCACCAAGAGACCAAGCTTGAAGGAGCACTGCTGATCTTCAAGGTG 360
Db 301 GGCATCTTACCGCCACCAAGAGACCAAGCTTGAAGGAGCACTGCTGATCTTCAAGGTG 360
QY 361 AAGTGCAAGGACCAACTTCCCGCGAGCGCCCGCTGATGAGAGCAAGAGCGGCGGC 420
Db 361 AAGTGCAAGGACCAACTTCCCGCGAGCGCCCGCTGATGAGAGCAAGAGCGGCGGC 420
QY 421 TGGAGCCCAAGCAGAGGTGTGTACCCCGAGAGAGCGGCTGTGTGTGTGTGTGTGT 480
Db 421 TGGAGCCCAAGCAGAGGTGTGTACCCCGAGAGAGCGGCTGTGTGTGTGTGTGTGT 480
QY 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATCTGCCACCACTACCAAGCTTACCGAG 540
Db 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATCTGCCACCACTACCAAGCTTACCGAG 540
QY 541 AAGAAGCGGTGCGGCGCTGACATGCGCGGCTTCCACTTCCAGCACTCCGGCTCCAG 600
Db 541 AAGAAGCGGTGCGGCGCTGACATGCGCGGCTTCCACTTCCAGCACTCCGGCTCCAG 600
QY 601 ATGCTGCGAAG 660
Db 601 ATGCTGCGAAG 660
QY 661 AGCGACCTGCCGAGAGAGGCAACTGA 687
Db 661 AGCGACCTGCCGAGAGAGGCAACTGA 687

RESULT 2
LOCUS AX527900 687 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 13 from Patent W00230965.
ACCESSION AX527900
VERSION AX527900.1 GI:25172349
KEYWORDS
SOURCE Heteractis criepa (leathery sea anemone)
ORGANISM Heteractis criepa
Stichodactylidae; Heteractis.

REFERENCE
1

AUTHORS Lukyanov,S.A., Fradkov,A.F., Lukyanov,K.A. and Gurskaya,N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 13 18-APR-2002;
Clontech Laboratories Inc. (US)
FEATURES
source Location/Qualifiers
1..687
/organism="Heteractis criepa"
/mol_type="unassigned DNA"
/db_xref="taxon:15771"

ORIGIN

Query Match 97.7%; Score 671; DB 6; Length 687;
Best Local Similarity 98.5%; Pred. No. 2.5e-77;
Matches 677; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGTGAGCGGCTCTGTAAGAGAGATATGCGCATCAAGTGTACATGAGGGACCGTG 60
Db 1 ATGTGAGCGGCTCTGTAAGAGAGATATGCGCATCAAGTGTACATGAGGGACCGTG 60
QY 61 AACGGCACTACTTCAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
Db 61 AACGGCACTACTTCAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
QY 121 AGCATGAAATCCAGTGAACGAGGGGCGCCCTTGCTTGCTTGCTTGCTTGCTTGCT 180
Db 121 AGCATGAAATCCAGTGAACGAGGGGCGCCCTTGCTTGCTTGCTTGCTTGCTTGCT 180
QY 181 CCTGTGCGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db 181 CCTGTGCGAGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 241 TTCAAGCAGAGCTTCCCGAGGGGCTTCACTGAGGAGAGAACCAACCACTGAGAGGAG 300
Db 241 TTCAAGCAGAGCTTCCCGAGGGGCTTCACTGAGGAGAGAACCAACCACTGAGAGGAG 300
QY 301 GGCATCTTACCGCCACCAAGAGACCAAGCTTGAAGGAGCACTGCTGATCTTCAAGGTG 360
Db 301 GGCATCTTACCGCCACCAAGAGACCAAGCTTGAAGGAGCACTGCTGATCTTCAAGGTG 360
QY 361 AAGTGCAAGGACCAACTTCCCGCGAGCGCCCGCTGATGAGAGCAAGAGCGGCGGC 420
Db 361 AAGTGCAAGGACCAACTTCCCGCGAGCGCCCGCTGATGAGAGCAAGAGCGGCGGC 420
QY 421 TGGAGCCCAAGCAGAGGTGTGTACCCCGAGAGAGCGGCTGTGTGTGTGTGTGTGT 480
Db 421 TGGAGCCCAAGCAGAGGTGTGTACCCCGAGAGAGCGGCTGTGTGTGTGTGTGTGT 480
QY 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATCTGCCACCACTACCAAGCTTACCGAG 540
Db 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATCTGCCACCACTACCAAGCTTACCGAG 540
QY 541 AAGAAGCGGTGCGGCGCTGACATGCGCGGCTTCCACTTCCAGCACTCCGGCTCCAG 600
Db 541 AAGAAGCGGTGCGGCGCTGACATGCGCGGCTTCCACTTCCAGCACTCCGGCTCCAG 600
QY 601 ATGCTGCGAAG 660
Db 601 ATGCTGCGAAG 660
QY 661 AGCGACCTGCCGAGAGAGGCAACTGA 687
Db 661 AGCGACCTGCCGAGAGAGGCAACTGA 687

RESULT 3
LOCUS AX527902 1396 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 15 from Patent W00230965.
ACCESSION AX527902
VERSION AX527902.1 GI:25172350
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct

DEFINITION Sequence 7 from Patent WO0230965.
 AX527894
 VERSION AX527894.1 GI:25172346
 KEYWORDS
 SOURCE
 ORGANISM
 Heteractis crispata (leathery sea anemone)
 Heteractis crispata
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Stichodactylidae; Heteractis.
 REFERENCE
 1 Lukyanov, S.A., Fradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 7 18-APR-2002;
 JOURNAL Clontech Laboratories Inc. (US)
 Location/Qualifiers
 FEATURES
 source 1..681
 /organism="Heteractis crispata"
 /mol_type="unassigned DNA"
 /db_xref="taxon:175771"

Query Match 64.2%; Score 441.2; DB 6; Length 681;
 Best Local Similarity 78.2%; Pred. No. 8.9e-48;
 Matches 530; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

10 GGCCTGCTGAAGAGATATGGCATCAAGATGATGAGGGGCAACGGTGAAGGGCCAC 69
 4 GGTGTTGTGAAGAAAGTATGGCATCAAGATGATGAGAGGCAACGGTTAATGGCAT 63
 70 TACTTCAAGTGCAGAGGGCGAGCGCAACCCCTTGGCCGGCAACCCAGAGATGAGA 129
 64 TATTCAAGTGAAG 123
 130 ATCCAGCTGACCGAGGGGCGCCCTGCTGCTTGCCTTCAGACATCCCTGAGCCCTGCTGC 189
 124 ATTCAATGTCACGAAAGGGGCTCCATTACATTTTCTTGACATTTTGGCAACGATTTGT 183
 190 GAGTACGGAG 249
 184 GAGTACGGAG 243
 250 AGCTTCCCGAGGGCTTCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
 244 TCTTTCCCTGAAGGCTTAACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 303
 310 ACCGCCACAG 369
 304 ACTGCTCATCAG 363
 370 GGGACCAACTTCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
 364 GGTACCAATTTCTCGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
 430 AGCAGCAGAGTGTGATACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
 424 AGCAGCAGAGTGTGATACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483
 490 AAGGTGGGCGAGCCGGGACCTGATTCGACCACTACCAAGAGAGAGAGAGAGAGAG 549
 484 AAGGTGGGCGAGCCGGGACCTGATTCGACCACTACCAAGAGAGAGAGAGAGAGAG 543
 550 GTGCGGCGCTGACCATGCGCGGCTTCACTTCAACGAGATCCGGCTCAAGATGCTGG 609
 544 GTCCGTGCTTGAACAATGCGAGATTTCAATTTTAAAGAGATCCGGCTTCAAGATGCTAGG 603
 610 AAGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
 604 AAGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663
 670 CCGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 687
 664 CTTGAAGAGAGAGAGATTTGA 681

RESULT 6
 AX527896
 LOCUS 681 bp DNA linear PAT 21-NOV-2002
 DEFINITION
 AX527896
 VERSION AX527896.1 GI:25172347
 KEYWORDS
 SOURCE
 ORGANISM
 Heteractis crispata (leathery sea anemone)
 Heteractis crispata
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Stichodactylidae; Heteractis.
 REFERENCE
 1 Lukyanov, S.A., Fradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 9 18-APR-2002;
 JOURNAL Clontech Laboratories Inc. (US)
 Location/Qualifiers
 FEATURES
 source 1..681
 /organism="Heteractis crispata"
 /mol_type="unassigned DNA"
 /db_xref="taxon:175771"

Query Match 64.0%; Score 439.6; DB 6; Length 681;
 Best Local Similarity 78.0%; Pred. No. 1.4e-47;
 Matches 529; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

10 GGCCTGCTGAAGAGATATGGCATCAAGATGATGAGGGGCAACGGTGAAGGGCCAC 69
 4 GGTGTTGTGAAGAAAGTATGGCATCAAGATGATGAGAGGCAACGGTTAATGGCAT 63
 70 TACTTCAAGTGCAGAGGGCGAGCGCAACCCCTTGGCCGGCAACCCAGAGATGAGA 129
 64 TATTCAAGTGAAG 123
 130 ATCCAGCTGACCGAGGGGCGCCCTGCTGCTTGCCTTCAGACATCCCTGAGCCCTGCTGC 189
 124 ATTCAATGTCACGAAAGGGGCTCCATTACATTTTCTTGACATTTTGGCAACGATTTGT 183
 190 GAGTACGGAG 249
 184 GAGTACGGAG 243
 250 AGCTTCCCGAGGGCTTCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
 244 TCTTTCCCTGAAGGCTTAACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 303
 310 ACCGCCACAG 369
 304 ACTGCTCATCAG 363
 370 GGGACCAACTTCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
 364 GGTACCAATTTCTCGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
 430 AGCAGCAGAGTGTGATACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 489
 424 AGCAGCAGAGTGTGATACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483
 490 AAGGTGGGCGAGCCGGGACCTGATTCGACCACTACCAAGAGAGAGAGAGAGAGAG 549
 484 AAGGTGGGCGAGCCGGGACCTGATTCGACCACTACCAAGAGAGAGAGAGAGAGAG 543
 550 GTGCGGCGCTGACCATGCGCGGCTTCACTTCAACGAGATCCGGCTCAAGATGCTGG 609
 544 GTCCGTGCTTGAACAATGCGAGATTTCAATTTTAAAGAGATCCGGCTTCAAGATGCTAGG 603
 610 AAGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 669
 604 AAGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663
 670 CCGAGAGAGAGAGAGATCTTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 687

Db 664 CTTGAAAAGCAATTGA 681

RESULT 7
AX527910
LOCUS AX527910 684 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 23 from Patent WO0230965.
ACCESSION AX527910
VERSION AX527910.1 GI:25172356
KEYWORDS
SOURCE
ORGANISM
Heteractis cripsa (leathery sea anemone)
Heteractis cripsa
Eukaryote; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
Stichodactylidae; Heteractis.

REFERENCE
AUTHORS Lukyanov, S.A., Pradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 23 18-APR-2002;
Clontech Laboratories Inc. (US)
LOCATION/Qualifiers

FEATURES
source
1. .684
/organism="Heteractis cripsa"
/mol_type="unassigned DNA"
/db_xref="taxon:175771"

ORIGIN
Query Match 63.8%; Score 438.2; DB 6; Length 684;
Best Local Similarity 77.6%; Pred. No. 2,2e-47;
Matches 530; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 5 TGAAGCGCCCTGCTGAGAGAGATATGCGATCAAGATGATCATGAGAGGACCGGTGAACG 64
Db 2 TGGCTGGTTTGTGTAAGAAAGTATGCGCATCAAGATGATCATGAGAGGACCGGTATATG 61

QY 65 GCCACTACTTCAAGTGGAGGGGCGAGCGGCAACCCCTTGCCTGGACCCAGCA 124
Db 62 GCATTAATTTCAAGTGTGAAGAGAGAGAGAGCGGCAACCATTTACAGTACGAGAGCA 121

QY 125 TGAAGATTCACGTACGAGAGGCGCCCTGCTGCGCTTGCATCTGAGATCTGGACCCCT 184
Db 122 TGAGATTCATGTACCGAAGGGGCTCATTAACATTTGCTTGCATTTTGGACCCGT 181

QY 185 GCTGCGAGTACGCGAGAGAGACTTCTGTGACCAACCGCGAGATCCCGACTTCTTCA 244
Db 182 GTTGTAGTACGCGAGAGAGACTTGTTCACCATTAACGAGAGATCCCGATTTCTTCA 241

QY 245 AGCAGAGCTTCCCGAGGGCTTCACTTGGAGAGACCACTTACGAGAGCGCGGCA 304
Db 242 AGCAGTCTTTCCTTGAAGGGCTTTACTTGGAGAGAGACCACTTATGAAGATGGAGCA 301

QY 305 TCCGTGACCGCCACAGAGACCAAGCTGGAGGGCACTGCTGATCTACAAAGGTGAAG 364
Db 302 TTCTTACTGCTCATCAGAGACCAAGCTGGAGGGCACTGCTTATATACAAAGGTGAAG 361

QY 365 TGCACGCGACCACTTCCCGCGCGACCGCCCGTGTGAAGAACAGAGCGCGGCTGGG 424
Db 362 TCTGTGTACCAATTTTCTGTGTATGCGCCCGTGAAGAACAAATCAGAGAGATGGG 421

QY 425 AGCCAGCAGCAGAGGTGTGTATCCCGAGAACCGCGTGTGGCGCGGAAAGTATGG 484
Db 422 AGCCAGCAGCAGAGGTGTGTATCCAGAGATGTGTCTGTGTGACGTAAATGTATGG 481

QY 485 CCGTGAAGGTGGGGGACCGGCACTGATCTGCCACCACTACACAGCTACCGAGCAAGA 544
Db 482 CCTTAAAGTGGGTGTGTGTGTGTATCTGCCATCTCTATATCTTTTACAGGTCCAGA 541

QY 545 AGGCGGTGCGCGCCCTGACCATGCGCGCTTCCATTACCGACATCCGCTCCAGATGC 604
Db 542 AAGCAGTCCGTGCTTACATGCGAGATTTTCAATTTTACAGACATCCGCTTCAAGTGC 601

QY 605 TGCAGAGAGAGAGAGAGATCTTCAAGCTGTACGAGCGCAGCGTGGCCCGGTACAGCG 664
Db 602 CGAGAGAGAGAGAGAGAGATCTTCAAGCTGTACGAGAGATCTGTGGCTTATGAGTGC 661

QY 665 ACCTGCCCGAGAGGCCAACTGA 687
Db 662 ATCTTCCTGAAAAGCAATTGA 684

RESULT 8
AX527892
LOCUS AX527892 684 bp DNA linear PAT 21-NOV-2002
DEFINITION Sequence 5 from Patent WO0230965.
ACCESSION AX527892
VERSION AX527892.1 GI:25172345
KEYWORDS
SOURCE
ORGANISM
Heteractis cripsa (leathery sea anemone)
Heteractis cripsa
Eukaryote; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;
Stichodactylidae; Heteractis.

REFERENCE
AUTHORS Lukyanov, S.A., Pradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.
TITLE Nucleic acids encoding stichodactylidae chromoproteins
JOURNAL Patent: WO 0230965-A 5 18-APR-2002;
Clontech Laboratories Inc. (US)
LOCATION/Qualifiers

FEATURES
source
1. .684
/organism="Heteractis cripsa"
/mol_type="unassigned DNA"
/db_xref="taxon:175771"

ORIGIN
Query Match 63.6%; Score 436.6; DB 6; Length 684;
Best Local Similarity 77.5%; Pred. No. 3.5e-47;
Matches 529; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 5 TGAAGCGCCCTGCTGAGAGAGATATGCGATCAAGATGATCATGAGAGGACCGGTGAACG 64
Db 2 TGGCTGGTTTGTGTAAGAAAGTATGCGCATCAAGATGATCATGAGAGGACCGGTATATG 61

QY 65 GCCACTACTTCAAGTGGAGGGGCGAGCGGCAACCCCTTGCCTGGACCCAGCA 124
Db 62 GCATTAATTTCAAGTGTGAAGAGAGAGAGAGCGGCAACCATTTACAGTACGAGAGCA 121

QY 125 TGAAGATTCACGTACCGAGAGGCGCCCTGCTGCGCTTGCATCTGAGATCTGGACCCCT 184
Db 122 TGAGATTCATGTACCGAAGGGGCTCATTAACATTTGCTTGCATTTTGGACCCGT 181

QY 185 GCTGCGAGTACGCGACAGAGACTTGTGTACCAACCGCGAGATCCCGACTTCTTCA 244
Db 182 GTTGTAGTACGCGACAGAGACTTGTGTCCACATACGCGAGATTTCCGATTTCTTCA 241

QY 245 AGCAGAGCTTCCCGAGGGCTTCACTTGGAGAGAACCACTTACGAGAGCGCGGCA 304
Db 242 AGCAGTCTTTCCTTGAAGGGCTTACTTGGAGAGAACCACTTATGAAGATGGAGCA 301

QY 305 TCCGTGACCGCCACAGAGACCAAGCTGGAGGGCACTGCTGATCTACAAAGGTGAAG 364
Db 302 TTCTTACTGCTCATCAGAGACCAAGCTGGAGGGCACTGCTTATATACAAAGGTGAAG 361

QY 365 TGCACGCGACCACTTCCCGCGCGACCGCCCGTGTGAAGAACAGAGCGCGGCTGGG 424
Db 362 TCTGTGTACCAATTTTCTGTGTATGCGCCCGTGAAGAACAAATCAGAGAGATGGG 421

QY 425 AGCCAGCAGCAGAGGTGTGTATCCCGAGAACCGGCTGTGTGCGCGGAAAGTATGG 484
Db 422 AGCCAGCAGCAGAGGTGTGTATCCAGAGATGTGTCTGTGTGACGTAAATGTATGG 481

QY 485 CCGTGAAGGTGGGGGACCGGCACTGATCTGCCACCACTACACAGCTACCGAGCAAGA 544
Db 482 CCTTAAAGTGGGTGTGTGTGTGTATCTGCCATCTCTATATCTTTTACAGGTCCAGA 541

QY 545 AGGCGGTGCGCGCCCTGACCATGCGCGCTTCCATTACCGACATCCGCTCCAGATGC 604
Db 542 AAGCAGTCCGTGCTTACATGCGAGATTTTCAATTTTACAGACATCCGCTTCAAGTGC 601

QY	603	TCCGGAAGAAAGAGAGACTTCTGAGCTGTACAGAGCCAGCTGGCCCGCGTACAGCG	664
Db	602	CGAGGAAACGAAAGACAGCTACTTTGAACTGTACGAGCATCTGTGCTAGTACAGTG	661
QY	665	ACCTGCCCGAGAAAGCCCACTGA	687
Db	662	ATCTTCTGTAAAGCAAAATGA	684

RESULT 9

LOCUS	AF363776	908 bp	mRNA	linear	INV 05-NOV-2001
DEFINITION	Heterocytis crispa GFP-like chromoprotein mRNA, complete cds.				
ACCESSION	AF363776				
VERSION	AF363776.1	GI:16660115			

ORGANI

SOURCE ORGANISM	REFERENCE
<i>Heteractis crispata</i> (leathery sea anemone)	
<i>Heteractis crispata</i>	
<i>Eukaryota</i> ; Metazoa; Chordaria; Anthozoa; Zoantharia; Actinoptaria	
Stichodactylidae, Heteractis.	
1 (bases 1 to 508).	
Archives	

AUTHOR:

TITLE Martynov, V.I., Yanushevich, Y.G., Lukyanov, K.A. and Lukyanov, S.A. GPP-like chromoproteins as a source of far-red fluorescent proteins

JOURNAL FEBS Lett. 507 (1), 16-20 (2001)

MENTION 31536606

REFERENCES

NUMBER	TITLE
0018454	Direct Submission
0018455	Gurskaya, N.G., Lukyanov, K.A., Labas, Y.A. and Lukyanov, S.A.

FEATURES

FEATURES
Mikiukho-Maklaya 16/10, Moscow 117997, Russia
location/Qualifiers

```

/organism="Heteractis crispa"
/mol_type="mrna"
/db_xref="taxon:175771"
CDS
77..760

```

```

/product="GFP-like chromoprotein"
/protein_id="AAL27538.1"
/db_xref="GI:1660116"
/translation="MAGLLKESMRIRMTVEGVANGHYKPCGEGDDNPPTGOSMRIRI
VTEGAPLPAPFDILAPCCGEYGTFFVHHAETLPDFKOSPGTEGRTVATYDGGII
THAOTSLEGNCLIKYKVGATFPADGVKMSKATPCTGVVY PENVGTCGNNV
ALKVDRLILCHLYTSRSKKAVALRMVGGFTDRLQMPRKCKDEYELVLAQVAF
VSDLPKPKAN"

```

ORIGIN

Query Match	63.6%	Score 436.6	DB 3	Length 908
Best Local Similarity	77.5%	Pred. No. 3.3e-47		
Matches 529; Conservative	0	Mismatches 154	Indels 0	Gaps 0

2Y	5	TGAGGGGCTGCTGCTGAAGAGAGATTAAGCCATCAAGATTATACATGAGGGGCAACCTGACG	64
3b	75	TGGCTGGTTGTGTGAAGAAATATATGCCATCAAGATGATACATGAGAGGCAAGGTTAAAG	13
2Y	65	GGCACTACTTCAGTGTGAGGCGGAGGCGAGGCGCAACCCCTGCGCGGACCCAGAGCA	12
3b	138	GGCATATTATTCAGTGTGAGAGAGAGGAGACCGCAACCCATTACAGGTATGCGAGAGCA	19
2Y	125	TGAGATATCCAGTGAACCGAGGGGCGCCCGCTGGCTTGCGCTTTGGATCATCTGGCCCT	18
3b	198	TGAGGATTCATGTCAACCGAAGGGGCTCATATTACATTGCGCTTGACATTTTGGCAACGT	25
2Y	185	GCTCGAGTACGGGCAAGAGACCTTCGTGACACACACGGCGAGATCCCGACCTTTTCA	24
3b	258	GTTGTGAGTACGGGCAAGAGACCTTTGTCCACATACGGCAGAAATTTCCGATTTCTTCA	31
2Y	245	AGCGACACTCCCCGAGGGCTTCACTCTGGAGAGAACCAACCACTTACGAGAGCGCGCA	30
3b	318	AGCAGCTTTTCCCTGAAAGCTTTACTTTGGAGAAAGAACCAACCTATGAAGATGAGAGCA	37

QY 305 TCTGTACGCCCAACAGAACACAGCGCTGTGAGGGCAATGCTGTATCTTACAGGTGAAAG 364

Db 378 TTCTTACGTCTCATCAGAGACAAAGCTGTGAGGGAACTGCTTATATACAAAGTGAAAG 437

QY 365 TGTAGCGCACCACTTCTCCCGCCGACGCGCCCTGTATGAAAGAACAGAGCGCGGCTGGG 424

Db 438 TCTTGTGTACCAATTTCTGCTGTATGCGCCCGGTGATGAAGAACAAATATCGAGAGATGGG 497

QY 425 AGCCAGACACCGAGTGTGTATCCCGAGAAACGCGCTGTGCGCGCGGAAACGTATGG 484

Db 498 AGCATGTGACTGAGTGTGTATTCAGAGAAATGTGTGTCGTGTGTGAGACGTATATGTATGG 557

QY 485 CCTTAAGTGTGGGAGACCGGGACCTGATCTGTGCCACATTCACAGCTAACCGAGACAAG 544

Db 558 CCTTAAAGTGTGGTATCGTCTGTTAACTGTGATCTGTGCCATCTCTATCTTTCACAGGTCCAAAG 617

QY 545 AAGCGGTGCGGCGCTGTACCATATGCCGGCTTCCACTTCAACGACATCCGCTCCAGATGC 604

Db 618 AAGCAGTCCGTCCTTGAACATGACGAGGATTTATTTTACACATCTCGCTTCAGATGG 677

QY 605 TCCGAGAGAAAGAGACGATCTTCGAGCTGTACAGAGCCAGCGTGGCCCGGTATACAGG 664

Db 678 CGAGGAAAAAGAAAGACGATCTTTGAACTGTACGACACATCTGTGTAGTACGATG 737

QY 665 AACTGTCCCGAGAGAGCCACTGA 687

Db 738 ATCTTCTCTGAAAAAGCAATTTGA 760

RESULT 1

RESULT 10	AX527888		
LOCUS	AX527888	910 bp	DNA
DEFINITION	Sequence 1 from Patent WO0230965.		linear
ACCESSION	AY507888		

KEYWORDS

LEAFY AN KEYWORDS	AN:27/000.1	GI:251/2334
SOURCE	Heteractis	crispa (Lea
ORGANISM	Heteractis	crispa

REFERENCE

REFERENCE

AUTHORS

TITLES

1 Stichodactylidae, Heteractis.

Lukyanov, S. A., Fradkov, A. F., Lukyanov, K. A. and Gurskaya, N. G.

Nucleic acids associated with the

END PAGE

FEATURES	Clontech Laboratories Inc. (US)
source	Location/Qualifiers 1..910

```

ORIGIN
        /organism="Heteractis cripea"
        /mol_type="unassigned DNA"
        /db_xref="taxon:175771"

Query Match      63.3%  Score 435;  DB 6;  Length 910;
Best Match Similarity  77.3%  Pred. No.5.3e-47;
Matches 528;  Conservative 0;  Mismatches 155;  Indels 0;  Gaps 0;

```

QY 5 TGAACGGCTCTGTAAGAGAGATATGGCATCAAGATGTATCATGAGAGGACCGTGAACG 64
Db 83 TGCCTGATTTGTTGAAGAAAGATATGGCATCAAGATGTATCATGAGAGGACCGTTAATG 142
QY 65 GCCACTACTTCAAGTGCAGAGGCGAGAGGCGAACCCCTTGGCCGGAGCCGAGAGA 124
Db 143 GCCATATTATTCAAAGTGTGAAGAGAGAGAGACGGCAACCTTACAGAGTACCGAAGCA 202
QY 125 TGAAGATTCACGTGAACGAGGCGCCCTTGCCTTGCATCTGGCCCTT 184
Db 203 TGAAGATTATGTACACGAAAGGCGCTCCATTTCCATTTGCTTCGACATTTTGGACCGT 262
QY 185 GCTGCGAGTACGCGAGCAGGACCTTGTGTCAACAACGCGCGAGATCCCGACTTTCTTA 244
Db 263 GTGTGTAGTACGGCAGCAGGACCTTTGTGTCAACATACGCGAGAGATTTCCGATTTCTTCA 322

QY 245 AGCAGAGCTTCCCGAGGCTTCACTTGAGAGAACACCACTTACGAGAGCGCGCA 304
 |||||
 DB 323 AGCAGCTCTTCCCTGAAAGGCTTATCTTGGAAGAACACCACTTATGAGATGAGGCA 382
 |||||
 QY 305 TCCTGACCGGCCACAGAGACCAAGCCTTGAGAGGCACTGCTGATCTTAAGGTGAAG 364
 |||||
 DB 383 TTCTTACTGTCTACAGAGACCAAGCTTGAGAGGAACTGCTTATATACAAAGTGAAG 442
 |||||
 QY 365 TGCACGGCACCACTTCCCGCCGACCGGCTGATGAAAGAACAGAGCGGCTGGG 424
 |||||
 DB 443 TCCTGTGTACCAATTTTCTGCTGATGAGCCCGATGAGAAACAAATCAGAGAGATGG 502
 |||||
 QY 425 AGCCACGACCGAGGTGTGTACCCCGAGAACCGGCTGTGCGCGGAAAGTGTGG 484
 |||||
 DB 503 AGCCATGTCACTGAGTGTGTATCAAGAAATGTGTCTGTGTGACGTAAATGTGATGG 562
 |||||
 QY 485 CCTGAAAGTGGGCGACCGGCACTGTATCTGCCACCACTACAGCTACCGAGAGCA 544
 |||||
 DB 563 CCTTAAAGTGGGATGTGTGTGATGTGCACATCTTATCTTACAGAGTCCAAAG 622
 |||||
 QY 545 AGGCGGTGCGGCGCTGACCATGCGCGGCTTCCACTTCAACGACATCGGCTCCAGATGC 604
 |||||
 DB 623 AAGCAGTCCGTGCTTGAACAATGCGAGATTTCAATTTACAGACATCCGCTTCAGATGC 682
 |||||
 QY 605 TGCAGAAAGAAAGACGAGTACTTTCAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCG 664
 |||||
 DB 683 CGAGAAACGAAAGACGAGTACTTGAATCTGTACGAAAGCATCTGTGGTAAAGTACAGTG 742
 |||||
 QY 665 ACCTGCGCGGAGAGGCCAATCTGA 687
 |||||
 DB 743 ATCTTCTGAAAAAGCAAAATTGA 765
 |||||

RESULT 11

AX527914

LOCUS AX527914 910 bp DNA linear PAT 21-NOV-2002
 DEFINITION Sequence 27 from Patent WO0230965.
 AX527914
 ACCESSION AX527914.1 GI:25172358
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Heteractis cripsa (leathery sea anemone)
 Heteractis cripsa
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Stichodactylidae; Heteractis.

REFERENCE 1
 LUKYANOV, S.A., FRADKOV, A.F., LUKYANOV, K.A. and GURSKAYA, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 27 18-APR-2002;
 Clontech Laboratories Inc. (US)

FEATURES
 source
 1..910
 /organism="Heteractis cripsa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:175771"

ORIGIN

Query Match 63.3%; Score 435; DB 6; Length 910;
 Best Local Similarity 77.3%; Pred. No. 5.3e-47;
 Matches 528; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
 QY 5 TGAAGCGGCTGTGAGAGAGATATGCGATCAAGATGTACATGAGAGGACCGTGAACG 64
 |||||
 DB 83 TGGCTGTGTTTGTGAAGAAAGTATGCGATCAAGATGTACATGAGAGGACCGTTTAATG 142
 |||||
 QY 65 GCCACTACTTCAAGTGTGAGAGGCGAGGCGCAAGGCAACCCCTTGCGCGGACCCAGAGCA 124
 |||||
 DB 143 GCCATTATTTTCAAGTGTGAAGAGAGAGGAGACGCGCAACCACTTATCAAGTATCGAGAGCA 202
 |||||
 QY 125 TGAAGATCCAGTGAAGAGGCGGCGCCCGCTGCGCTTGCCTTGAATCTGGCCCCCT 184
 |||||
 DB 203 TGAAGATTCATGTACAGAAAGGCGCTCAATTCATTTGCTTGAATTTTGGCACCGT 262
 |||||
 QY 185 GCTGCGAGTACGCGACGAGACCTTCTGTGACCAACCGCGCGAGATCCCGAATTTCTTCA 244
 |||||

DB 263 GTTGTAGTACGCGACAGACCTTGTCCACATACGCGAGGATTCGGAATTTCTTCA 322
 |||||
 QY 245 AGCAGAGCTTCCCGAGGCTTCACTTGAGAGAACACCACTTACGAGAGCGCGCA 304
 |||||
 DB 323 AGCAGCTCTTCCCTGAAAGGCTTATCTTGGAAGAACACCACTTATGAGATGAGGCA 382
 |||||
 QY 305 TCCTGACCGGCCACAGAGACCAAGCCTTGAGAGGCACTGCTGATCTTAAGGTGAAG 364
 |||||
 DB 383 TTCTTACTGTCTACAGAGACCAAGCTTGAGAGGAACTGCTTATATACAAAGTGAAG 442
 |||||
 QY 365 TGCACGGCACCACTTCCCGCCGACCGGCTGATGAAAGAACAGAGCGGCTGGG 424
 |||||
 DB 443 TCCTGTGTACCAATTTTCTGCTGATGAGCCCGATGAGAAACAAATCAGAGAGATGG 502
 |||||
 QY 425 AGCCACGACCGAGGTGTGTACCCCGAGAACCGGCTGTGCGCGGAAAGTGTGG 484
 |||||
 DB 503 AGCCATGTCACTGAGTGTGTATCAAGAAATGTGTCTGTGTGACGTAAATGTGATGG 562
 |||||
 QY 485 CCTGAAAGTGGGCGACCGGCACTGTATCTGCCACCACTACAGCTACCGAGAGCA 544
 |||||
 DB 563 CCTTAAAGTGGGATGTGTGTGATGTGTGATGTGCACATCTTATCTTACAGAGTCCAAAG 622
 |||||
 QY 545 AGGCGGTGCGGCGCTGACCATGCGCGGCTTCCACTTCAACGACATCGGCTCCAGATGC 604
 |||||
 DB 623 AAGCAGTCCGTGCTTGAACAATGCGAGATTTCAATTTACAGACATCCGCTTCAGATGC 682
 |||||
 QY 605 TGCAGAAAGAAAGACGAGTACTTTCAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCG 664
 |||||
 DB 683 CGAGAAACGAAAGACGAGTACTTGAATCTGTACGAAAGCATCTGTGGTAAAGTACAGTG 742
 |||||
 QY 665 ACCTGCGCGGAGAGGCCAATCTGA 687
 |||||
 DB 743 ATCTTCTGAAAAAGCAAAATTGA 765
 |||||

RESULT 12

AX527890

LOCUS AX527890 908 bp DNA linear PAT 21-NOV-2002
 DEFINITION Sequence 3 from Patent WO0230965.
 AX527890
 ACCESSION AX527890.1 GI:25172344
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Heteractis cripsa (leathery sea anemone)
 Heteractis cripsa
 Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
 Stichodactylidae; Heteractis.

REFERENCE 1
 LUKYANOV, S.A., FRADKOV, A.F., LUKYANOV, K.A. and GURSKAYA, N.G.
 Nucleic acids encoding stichodactylidae chromoproteins
 Patent: WO 0230965-A 3 18-APR-2002;
 Clontech Laboratories Inc. (US)

FEATURES
 source
 1..908
 /organism="Heteractis cripsa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:175771"

ORIGIN

Query Match 62.9%; Score 431.8; DB 6; Length 908;
 Best Local Similarity 77.0%; Pred. No. 1.4e-46;
 Matches 526; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
 QY 5 TGAAGCGGCTGTGAGAGAGATATGCGATCAAGATGTACATGAGAGGACCGTGAACG 64
 |||||
 DB 78 TGGCTGTGTTTGTGAAGAAAGTATGCGATCAAGATGTACATGAGAGGACCGTTTAATG 137
 |||||
 QY 65 GCCACTACTTCAAGTGTGAGAGGCGAGGCGCAAGGCAACCCCTTGCGCGGACCCAGAGCA 124
 |||||
 DB 138 GCCATTATTTTCAAGTGTGAAGAGAGAGGAGACGCGCAACCACTTATCAAGTATCGAGAGCA 197
 |||||
 QY 125 TGAAGATCCAGTGAAGAGGCGGCGCCCGCTGCGCTTGCCTTGAATCTGGCCCCCT 184
 |||||

Db 198 TGAGGATTTCATGTCAACCGAAGGGGCTCCATTACATTTGCTTCGACATTTTGGACCCGT 257

Qy 185 GGTGGAGTATACGAGACAGAGACTTGTGTGCACACACCGGAGATCCCGACTTCTTCA 244

Db 258 GTTGTAGATACGAGACAGAGACTTGTGTGCACATACCGGAGATTTCCGATTTCTTCA 317

Qy 245 AGCAGAGCTTCCCGAGGGCTTCACTGAGAGAGAACCAACCTTACGAGACGGCGCA 304

Db 318 AGCAGCTTTTCCCTGAAAGGCTTACTTGGAAAGAACCAACCTTACGAGAGCA 377

Qy 305 TCTTGACCGCCACACAGACACACAGCTGAGAGGGACACTGCTGATCTTCAAGAGTAAAG 364

Db 378 TTCTTACTGTGCATACGAGACACAGCTGAGAGGGAACTGCTTATATACAGGTGAAAG 437

Qy 365 TGACAGGACCAACTTCCCGCGCAGACGGCCCTGTATGAAACAAAGCGCGCTGCA 424

Db 438 TCTTGTGACCAATTTTCTGCTGATGGCCCGATGAAACAAATCAGAAAGATGGG 497

Qy 425 AGCCACACACCGAGGTGTGTACCCCGAAGACGGCGTGTGCGCCGAAAGTATGG 484

Db 498 AGCATGCTACTGAGGTGTGTATCCAGATTAATGGTGTCTGTGAGCGTAAATGTATGG 557

Qy 485 CCCTGAAGGTGGCGACCGGCACTGATCTGCAACCACTACACAGCTACCGGAGCAAGA 544

Db 558 CCTTAAAGTCGTGTATCGTCTTGTATCTGCAATCTTATCTTTCAGGTCCAAAGA 617

Qy 545 AGGCGGTGCGCGCTTGAACATGCGCGCTTCCACTTACCGACATCCGGCTCCAGATGC 604

Db 618 AAGCAGTCGTGTCTTGAACATGCGAGATTTTCAATTTTACAGACATCGCGCTTCAGATGC 677

Qy 605 TGCGGAAAGAAAGAGACAGATCTTCCAGCTGTATCGAGGCCAGCGTGGCCGGTACAGCG 664

Db 678 CGAGAGAAAGAAAGAGACAGATCTTGAACGTATCGAAGCATCTGTGTAGTACAGTGG 727

Qy 665 ACTCTGCCGAGAAAGGCCAATCTGA 687

Db 738 ATCTCTCTGAAAAAGCAATTGA 760

RESULT 13

AX527912 680 bp DNA linear PAT 21-NOV-2002

LOCUS AX527912

DEFINITION Sequence 25 from Patent WO230965.

ACCESSION AX527912

VERSION AX527912.1 GI:25172357

KEYWORDS

SOURCE Heteractis crispa (leathery sea anemone)

ORGANISM Heteractis crispa

REFERENCE 1. Lukyanov, S.A., Fradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G. Nucleic acids encoding stichodactylidae chromoproteins Patent: WO 0230965-A 25 18-APR-2002; Clontech Laboratories Inc. (US)

AUTHORS Lukyanov, S.A., Fradkov, A.F., Lukyanov, K.A. and Gurskaya, N.G.

TITLE Nucleic acids encoding stichodactylidae chromoproteins

JOURNAL Patent: WO 0230965-A 25 18-APR-2002; Clontech Laboratories Inc. (US)

FEATURES

source location/Qualifiers

1..680

/organism="Heteractis crispa"

/mol_type="unassigned DNA"

/db_xref="taxon:175771"

ORIGIN

Query Match 62.5%; Score 429.2; DB 6; Length 680;

Best Local Similarity 78.0%; Pred. No. 3.1e-46;

Matches 529; Conservative 0; Mismatches 148; Indels 1; Gaps 1;

Qy 10 GGCCTGCTGAAGAGAGTATGCGCATCAAGATGTACATGAGAGGCAACCGTGAACGGCCAC 69

Db 4 GGTTTGTGAAAGAAAGATATGCGCATCAAGATGTACATGAGAGGCAACGGTTATGGCCAT 63

Qy 70 TACTTCAAGTGCAGAGGCGAGCGGCAACCGCTTCCGCGGCAACCGCAGAGCATGAGA 129

Db 64 TATTTCAAGTGTGAAAGAGAGGAGCGGCAACCGATTTGCAAGTACGAGACATGAGG 123

Qy 130 ATCCAGGTGACGAGAGGGGCCCCCGCTTGGCCCTTGGACATCTGTGCCCCCTGTCTGC 189

Db 124 ATTCATGTACACGAGAGGGGCTTCATTTACATTTGGCTTGGACATTTTGGACCCGTGTGT 183

Qy 190 GAGTACGACAGACGACCTTGTGTGACCAACCGCGAGATCCCGACTTCTTCAAGCAG 249

Db 184 GCGTACGAGACAGACCTTGTGTGACCAACCGAGAGATTTCCGATTTCTTCAAGCAG 243

Qy 250 AGCTTCCCGAGAGGCTTCACTGTGAGAGAACCAACCTTACGAGAGACGGCGCATCTTG 309

Db 244 TCTTTCCCTGAAGAGCTTTACTTGGGAAAGAACCAACCTTACGAGAGAGGCAATCTT 303

Qy 310 ACCGCCACACAGACACAGAGCTTGAAGGCACTGTGTGATCTTACAGAGTGAAGTGCAC 369

Db 304 ACTGCTCATACAGACACAGAGCTTGAAGGCACTGTGTGATCTTACAGAGTGAAGTGCAC 363

Qy 370 GGACCAACTTCCCGCGAGAGCGCCCGTGTATGAAACAAAGAGCGCGCTTGGAGCCC 429

Db 364 GGTACCAATTTTCTGCTGTATGGCCCGATGAAAGAAATTCAGAGAGATGGAGCCA 422

Qy 430 AGCATCGAGGTGTGTACCCCGAAGACGGCGTGTGTGCGCCGGAACGTATGGCCCTG 489

Db 423 AGCATGAGGTGTGTATCCAGAGAAATGGTGTCTGTGTGAGCGTAAATGTATGGCCCTT 482

Qy 490 AAGGTGGGCAACCGGCACTGATCTGCAACCACTACACAGCTACCGGAGCAAGAGGCC 549

Db 483 AAGGTGGGCAACCGGCACTGATCTGCAACCACTACACAGCTACCGGAGCAAGAGGCC 542

Qy 550 GTGCGGCGCTTGAACATGCGCGCTTCCACTTACCGACATCCGGCTCCAGATGCTGCGG 609

Db 543 GTCCGCGCTTGAACATGCGAGATTTTCAATTTTACAGACATCCGCTTCAGATGCTGAGG 602

Qy 610 AAGAAGAGAGAGAGTCTTCCAGCTGTATCGAGGCCAGCGTGGCCCGGTACAGCACTTG 669

Db 603 AAAAGAGAGAGAGTCTTGAACGTATCGAAGCATCTGTGTAGTACAGTGTACTT 662

Qy 670 CCGGAGAAAGGCCAATCTGA 687

Db 663 CCGTGAAGAAAGCAATTGA 680

RESULT 14

AY233273 699 bp mRNA linear SYN 02-MAR-2003

LOCUS AY233273

DEFINITION Synthetic construct kindling fluorescent protein mRNA, complete cds.

ACCESSION AY233273

VERSION AY233273.1 GI:28629492

KEYWORDS

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1. (bases 1 to 699) Artificial sequences.

AUTHORS Chudakov, D.M., Belousov, V.V., Zaralaky, A.G., Novoselov, V.V., Staroverov, D.B., Zorov, D.B., Lukyanov, S. and Lukyanov, K.A. Kindling fluorescent proteins for precise in vivo photolabeling Nat. Biotechnol. 21 (2), 191-194 (2003)

JOURNAL Nat. Biotechnol. 21 (2), 191-194 (2003)

FEATURES

source location/Qualifiers

1..699

/organism="synthetic construct"

/mol_type="mRNA"

/db_xref="taxon:32630"

1..699

/note="GFP-like protein; capable of photoconversion from nonfluorescent to red fluorescent form; KFP1; KFP-red"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2004, 11:17:23 ; Search time 82 Seconds
(without alignments)
4649.406 Million cell updates/sec

Title: US-09-976-673-11

Perfect score: 687
Sequence: 1 atggtgagcgcgtctgtaa.....tgccgagaggaacacatga 687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	253.2	36.9	720	4	US-09-839-650-1 Sequence 1, Appl
2	176.6	25.7	678	4	US-09-459-956-6 Sequence 6, Appl
3	153.8	22.4	801	4	US-09-459-956-7 Sequence 7, Appl
4	145.4	21.2	699	4	US-09-459-956-5 Sequence 5, Appl
5	121.6	17.7	1104	3	US-09-277-716-30 Sequence 30, Appl
6	121.6	17.7	1104	4	US-09-277-716-31 Sequence 31, Appl
7	121.6	17.7	1279	3	US-09-609-161B-31 Sequence 31, Appl
8	121.6	17.7	1279	4	US-09-609-161B-31 Sequence 14, Appl
9	110.6	16.1	720	4	US-09-316-919-15 Sequence 15, Appl
10	110.6	16.1	720	4	US-09-602-641-14 Sequence 14, Appl
11	110.6	16.1	690	4	US-09-459-956-2 Sequence 2, Appl
12	109.8	16.0	717	4	US-09-513-783A-43 Sequence 43, Appl
13	109	15.9	720	3	US-08-911-825-3 Sequence 3, Appl
14	109	15.9	720	3	US-08-974-737-3 Sequence 3, Appl
15	109	15.9	720	3	US-08-706-408-3 Sequence 3, Appl
16	109	15.9	720	3	US-09-094-359-9 Sequence 9, Appl
17	109	15.9	720	3	US-09-172-063-12 Sequence 12, Appl
18	109	15.9	720	3	US-09-172-063-17 Sequence 17, Appl
19	109	15.9	720	3	US-09-465-142-3 Sequence 3, Appl
20	109	15.9	720	4	US-09-316-919-13 Sequence 13, Appl
21	109	15.9	720	4	US-09-316-919-18 Sequence 18, Appl
22	109	15.9	720	4	US-09-575-847-3 Sequence 3, Appl
23	109	15.9	720	4	US-09-602-641-12 Sequence 12, Appl
24	109	15.9	720	4	US-09-602-641-17 Sequence 17, Appl
25	109	15.9	720	4	US-09-704-463-9 Sequence 9, Appl
26	109	15.9	768	3	US-09-172-063-28 Sequence 28, Appl

28	109	15.9	768	3	US-09-172-063-33 Sequence 33, Appl
29	109	15.9	768	4	US-09-602-641-28 Sequence 28, Appl
30	109	15.9	768	4	US-09-602-641-33 Sequence 33, Appl
31	109	15.9	972	3	US-09-172-063-30 Sequence 30, Appl
32	109	15.9	972	4	US-09-602-641-30 Sequence 30, Appl
33	109	15.9	1623	4	US-09-513-783A-33 Sequence 33, Appl
34	109	15.9	2439	4	US-09-513-783A-3 Sequence 3, Appl
35	109	15.9	2439	4	US-09-513-783A-5 Sequence 5, Appl
36	109	15.9	3171	4	US-09-513-783A-31 Sequence 31, Appl
37	109	15.9	4833	4	US-09-172-063-21 Sequence 21, Appl
38	107.8	15.7	720	3	US-09-172-063-18 Sequence 18, Appl
39	107.8	15.7	720	4	US-09-316-919-19 Sequence 19, Appl
40	107.8	15.7	720	4	US-09-602-641-18 Sequence 18, Appl
41	107.8	15.7	768	3	US-09-172-063-34 Sequence 34, Appl
42	107.8	15.7	768	4	US-09-602-641-34 Sequence 34, Appl
43	107.4	15.6	720	4	US-09-094-359-5 Sequence 5, Appl
44	107.4	15.6	720	4	US-09-704-463-5 Sequence 5, Appl
45	107.4	15.6	762	1	US-08-532-390-40 Sequence 40, Appl

ALIGNMENTS

```
RESULT 1
US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; TITLE OF INVENTION: Fluorescent Protein
; Patent No. 6645761
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc feature
; LOCATION: (1) - (720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Query Match      36.9%; Score 253.2; DB 4; Length 720;
Best Local Similarity 62.0%; Pred. No. 1.9e+43;
Matches 418; Conservative 0; Mismatches 253; Indels 3; Gaps 1;

2 TGGTGAACCGCTGCTGTAAGAGAGATGCGCATGAATGATGATGAGAGGACCGTGA 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 TGAAGAACACCTGCTGAGAGAGTGAAGTGAACCTGAGAGGACCTGAGAGGACCTGA 79
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 AGGGCACTACTTCAAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 AACACCACTGTTCCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 GCATGAGATCCAGTGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 181
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 TGGTGAAGATCGCGGTGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 199
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 CCTGCTGCGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 241
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 CCGCTTCAGTACGAGCAACGACCTTCAACCAAGTATCCCAACGACATCAACGACTACT 259
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 TCAAGAGAGCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 301
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 TATATCAGAGCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 319
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 GCATCTGACCCGACCAAGACACCAAGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 361
```

```

Db      320 GCGTGGTGAAGATCCGACGACATCACTGATCGAGGACAAAGTTGTGTACCGGTGG 379
Qy      362 AGGTGACGCGACCAATCTCCCGCGACGCGCGGTGATGAAGAAAGAGGGGCGTT 421
Db      380 AGTACAGGCGACCAATCTCCCGACGCGCGGTGATGAAGAAAGAGGGGCGTT 439
Qy      422 GGGAGCCGACGACGAGTGTGTACCCGAGAAAGCGGTGTGTGTGTGTGTGTGTGTGT 481
Db      440 TCGAGCCGACGTTGAGCGCATGTATGATGAAGAGGCGGTGTGTGTGTGTGTGTGT 499
Qy      482 TGGCCCTGAAAGTGGGAGACCGGACCTGTATGTGCAACCACTAACCGACTACCGAGCA 541
Db      500 TGGTGTACAGCTGAAGAGGAGCAAGTACTACAGTGTGCAACAGTAAAGACCTGATGAGA 559
Qy      542 AGAAGCCGTCGCGCGCGCTTGACCAATGCGCGCTTCACTTCAACGACATCCGCGTCA 601
Db      560 GCGAGGGCGTGTG--GTGAAGGAGTTCCCTCTCACTTCACTTCAACGACCGCGTGA 616
Qy      602 TGTGCGGAAAGAAAGAGAGTACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 661
Db      617 AGACCTACGTGAGGACGCGCGCTTGTGAGCAGCAGAGACCGCATCGCCAGATGA 676
Qy      662 GCGACCTGCGCGAG 675
Db      677 CCGACATCGCGAAG 690

```

```

RESULT 2
US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1999-05-08
; PRIOR FILING DATE: 08/765,860
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discoboma sp
US-09-459-956-6

```

```

Query Match      25.7%; Score 176.6; DB 4; Length 678;
Best Local Similarity 60.2%; Pred. No. 8.7e-28;
Matches 293; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

Qy      17 TGAAGGAGATATCCGATCAAGATGATCATGAGGACCGTGAACGCGCACTACTTCA 76
Db      23 TCAAGAGATTCATGAGGTTAAAGTTTCGATGAGAAAGAACGTCATATGGGACAGGTTTG 82
Qy      77 AGTCCGAGGCGGAGGCGACGCAACCCCTTCCGCGGACCCGAGACATAGGAATCCACG 136
Db      83 AATATGAAAGCGAAGAGAGGAGGAGGACATACGAAAGCACAATACCGTAAAGCTTAAAG 142
Qy      137 TGACGAGGCGCGCGCGCGCGCTTCCGCTTCCGACATCCGCGCGCGCGCGCGCGAG 196
Db      143 TAACCAAGGGGGGACCTTTGCGATTTGCTTGCGATATTTTGTGACCAATTTCAATG 202
Qy      197 GCGACGAGACCTTTCGTCACCAACCGCGGAGATCCCGACCTTTCGAGGAGAGCTTCC 256
Db      203 GAAGCAAGATATATGTCAGAGCAGCGCGCGACATACCAAGATTAATAAAGCTGTCAATTG 262

```

```

Qy      257 CCGAGGCGCTTCACTGAGAGAAACCAACCACTAGAGGACGCGCGCATCTGTACCGCCC 316
Db      263 CTGAAGAGATTTAAATGGGAAAGGTCATGAACCTTGAAGACGCGTCCGTCTACTTAA 322
Qy      317 ACCAGAGACCGACCTGAGAGGCGCAACGCTGTATCTTCAAGGTGAAGGTGACGCGACA 376
Db      323 CCGAGATTCGAGTTTGAAGATGAGCTTTTCACTTCAAGGTGAAGGTGACGCGACA 382
Qy      377 ACTTCCCGCGAGCGCGCGCGGTGATGAAGAAAGAGCGCGCGGTGAGAGCCAGACCG 436
Db      383 ACTTCTTCCGAGTGAACCTGTATCAAAAGAGCAATGCGCGTGGAGAGCCAGCACGTG 442
Qy      437 AGGTGTGTACCCGAGAAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 496
Db      443 AGGTTGTATCTTCGTGATGCGGTGTGAAGAGAGATTCATTAAGGCTCTGAAGCTGA 502
Qy      497 GCGACCG 503
Db      503 AAGACGG 509

```

```

RESULT 3
US-09-459-956-7
; Sequence 7, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1999-05-08
; PRIOR FILING DATE: 08/765,860
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Clavulaxia sp
US-09-459-956-7

```

```

Query Match      22.4%; Score 153.8; DB 4; Length 801;
Best Local Similarity 58.4%; Pred. No. 4.1e-23;
Matches 269; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Qy      10 GGCCTGTGAAGAGATATCCGATCAAGATGATCATGAGGACCGGTGAACGCGCAC 69
Db      130 GGTGTATTTAAACAGACATGAAGATTAAGCTGAAGTGAAGAAATGTAAACGCGCAT 189
Qy      70 TACTTCAAGTCCGAGGCGGAGGCGACGCAACCCCTTCCGCGGACCCAGACATGAGA 129
Db      190 GCTTTGTGATCGAAGAGAAAGAGAAAGGCTTACGATGGGACACACACTTTAAAC 249
Qy      130 ATCCAGTGAACGAGGCGCGCGCGCTTCCGCTTCCGCTTCCGATTCCTGTGCGCGCG 189
Db      250 CTGAAGATGAAGGAGGTCGCTGCTTTTCTTACGATTCCTGTGCGCGCGCGCGGTC 309
Qy      190 GAGTACGCGACAGACCTTTCGTGACCAACCGCGGAGATCCCGACCTTTCGAGCGAG 249
Db      310 CAGTACGGAACAGAGATTTGACAAATACCAAGCATATATGAGCTATTTCAAGCAG 369
Qy      250 AGCTTCCCGAGGCGCTTCACTGAGAGAAACCAACCTTACAGAGAGCGCGCATTCCTG 309
Db      370 TCGTTTCCGAGGATTTCTTGGAAAGAAACATGACTTTTGAAGACAAAGCATTTGTC 429
Qy      310 ACCCGCCACAGACACAGCTGAGAGGCAACTGCTGTATCTTCAAGGTGAAGGTGAC 369

```

Db 430 AAAGTAAAAAGTGCATTAAGCATGAGAGAAAGCTCCTTATCTATGAAATTCGTTTGAT 489
 370 GGCACCAACTTCCCCGCGCAGCGCCCCGTGATGAAACAAGAGCGCGCTGGAGCCC 429
Db 490 GGGATGAACCTTCTCCCAATGATGTCGGGTATAGCAGAAAAAACTTGAAGTGGAAACA 549
Dy 430 AGCAGCGAGGTGATACCCCGAAGAACGGCGTGTGGG 470
Db 550 TCACCTGAGATTATGTACGTGCGTGTGAGAGTGTGGTGG 590

```

RESULT 4
US-09-459-956--5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Telen, Roger Y.
; APPLICANT: Gonzalez, Iii, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
; US-09-459-956-5

```

Query Match	Best Local Match	Similarity	51.2%;	Score	145.4;	DB	4;	Length	699;
Matches	272;	Conservative	0;	Mismatches	211;	Indels	0;	Gaps	0
QY	14	TGCTGAAGAGAGACTATGCCATCTCAGATGTACATGAGAGGCAACCGTGACCGGCATCTACT	73						
Db	20	TGATCAAGAGAGAAATGTGATCGATCTTCATCTGTGAAGAAAGCTTCATATGGGCACTACT	79						
QY	74	TCAAGTGCAGGGCGAGGGCGGACGGCAACCCCTTGCGGGACCCAGACGATGAGAAATCC	133						
Db	80	TTGAAATTAAGGCAAGGAAAGAAAGACAGCCTTAATGAAAGGCAACCAATACGCTACGCTCG	139						
QY	134	ACGTGACCGAGGGGCGCCCTTGCCCTTGCCCTTGCAATCTTGCGCCCTGCTGCGAGT	193						
Db	140	AGGTTACCAAGGGGTGAGACCTGTGCCATTTGGTGCAATTTTGGCCCAATTTTCAGT	199						
QY	194	ACGGCAGCAGGACCTTCGTGCAACACCGCGGAGATCCCGACTTCTTAAGCAGAGCT	253						
Db	200	ATGGAATACAGGACTTTTGTTCACACACCTTGACAAATACATGATTAATTAAAGCTGTGCAT	259						
QY	254	TCCCGAGGGCTTCACTCGTGGAGAGAACCAACCACTTACAGAGACGGCGGCATCTGACCG	313						
Db	260	TTCCGGAGGGATATATCATGGGAAACGGTTCATGCATTTGAAAGACGGTGGCTTGTGTGTA	319						
QY	314	CCCAACAGAGACACACACTCGTGAAGGGCAATGCTGTATCTTACAGGTGAAGGTGACGGCA	373						
Db	320	TCACCAATGATATACAGTTTGAACAGGAACTGTTTCTACTACGACATCAAGTTCATCGGCT	379						
QY	374	CCAACTTCCCGCCGACGGCGCCCGTGATGAAAGACAAGACGGCGGCTGGAGCCCAACA	433						
Db	380	TGAATCTTCTCCAAATGAGCCGTTGTGCAAGAAAGACAATCGGCTGGGAAACGAGCA	439						
QY	434	CCGAGGTGTGTACCCCGAAGACGGGTGCTGTGGCGCCGAAAGCTATATGGCCTTGAAG	493						
Db	440	CTAGAGCTTTGATCTCTCGTGAATGGTGTGTATATGAGAAACATCCATCATCTCTGACAG	499						

```

RESULT 5
US-09-277-716-30
: Sequence 30. Application US/09277716A
: Patent No. 6232107
: GENERAL INFORMATION:
: APPLICANT: Bryan, Bruce
: APPLICANT: Szent-Gyorgyi, Christopher
: TITLE OR INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
: CURRENT APPLICATION NUMBER: US/09/277,716A
: CURRENT FILING DATE: 1999-03-26
: EARLIER APPLICATION NUMBER: 60/102,939
: EARLIER FILING DATE: 1998-10-01
: EARLIER APPLICATION NUMBER: 60/089,367
: EARLIER FILING DATE: 1998-06-15
: EARLIER APPLICATION NUMBER: 60/079,624
: EARLIER FILING DATE: 1998-03-27
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 30
: LENGTH: 1104
: TYPE: DNA
: ORGANISM: Ptilosarcus gurneyi
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (34)..(747)
: FEATURE:
: OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

```

Query Match	Similarity	17.7%;	Score 121.6;	DB 3;	Length 1104;
Best Local	Similarity	54.2%;	Pred. No. 1.7e-16;		
Matches	247;	Conservative	0;	Mismatches	209;
				Indels	0;
				Gaps	0;
QY	16	CTGAAGAGAGTATGCGCATCAAGATGTACATGAGAGGCACCGGTGAACGGCCACTACTTC	75		
Db	64	CTGAAGAAGATTTATGTGCGCAAAAGCTAAGCGTTGAAGGAATGCTGAACAATACAGTTTTT	123		
QY	76	AAGTGGAGGGCGCAGGGCGCAGCGCAACCCCTTGCCCGGACCCAGACGATGAGAAATCCAC	135		
Db	124	TCATGGAAGAGATTTGTGAAAAGGCAATGTATTTTGGAAACCAATGTAGTCAAAATCCGG	183		
QY	136	GTGACCGAGGGCGCCCCCTGCCCCCTTGCCCTTGCCTTGCATCTGGCCCCCTGCTCGAGTAC	195		
Db	184	GTTACAAAGGAGAGTCCGTTGCCATTCGCTTGCATATATGTTTCCATACCTTCCAAATAC	243		
QY	196	GGAGACAGACCTTTCGTGACCAACCGCGAGATCCCGACCTTCTTGAAGCAGAGCTTC	255		
Db	244	GGGAATGCACTTTCACGAATATCCAGAGACATTCGGACCTACTTGTTCATATATTC	303		
QY	256	CCGAGGGCTTCACTTGGAGAGAACCAACCACTTACGAGACGGCGGACATCTGACCGCC	315		
Db	304	CCGGCTGAGATTTTCTACGAAGAATCTAACGTTTGAAGATGGCCCCATTTGTGCATT	363		
QY	316	CACCAAGACACGAGCCCTGAGGGGCAACTGCGTATCTTACAAGGTGAAGTGAACGCGACCC	375		
Db	364	CGTTCAAGATTTAAGTTTAGAGTGAATTAAGTTCCACTTACCAAGGTGAAGTATAGAGCAAC	423		
QY	376	AACCTTCCCGCCGACGCGCCCGGTGATGAAGAACAGAGGGCGGCTGGAGCCCAAGACCC	435		
Db	424	GGTTTCCCTAGTAAACGAGACCCGTGATGCAAAAAGGCATCTCGGACGTGAGGCATGTTT	483		
QY	436	GAGTGCTGTACCCCGAAGAACGGCGGTGCTGTGGCG	471		
Db	484	GAGGTGCTTCACTGAACAGCGCGCTTCTGTGGCGC	519		

```
RESULT 6
US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 643682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (Insert A)
US-09-609-161B-30

Query Match      17.7%; Score 121.6; DB 4; Length 1104;
Best Local Similarity 54.2%; Pred. No. 1.7e-16;
Matches 247; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 16 CTGAAGAGAGATATCCGATCATGATGTATCATGAGAGGAGCCGCGCACTTCTTC 75
DB 64 CTGAAGAGATATATGTCGCGCAAAAGCTAGCGTTGAAGAAATGTGAACATCACTTTT 123
QY 76 AAGTGCAGAGGCGAGGCGGAGCGGCAACCCCTTCGCCGCGCACCCGAGATGAAATTCAC 135
DB 124 TCCATGGAAGAGATTGTGAAAAGGCAATGTAATTTTGAAGAACCAATTGATCAATCCG 183
QY 136 GTGACCGAGGCGCCCCCTTCGCTTCGCTTCGATCTTGAGCCCTTCGCGATGAC 195
DB 184 GTTACAAAGGAGAGGTCGTTGCGCATTCGCTTCGATTTGTTTCATAGCTTTCGAATAC 243
QY 196 GGCAGAGAGACCTTGTGCAACAACCGCGGAGATCCCGACTTCTTCAAGCAGACCTTC 255
DB 244 GGGAAATCCGACTTTCAGAAATACCGAGACGACATTCGCGACTCTTTGTTCATCATTC 303
QY 256 CCCGAGGCTTCACTGCGAGAGAACCACTTACGAGACGCGCGGCAATCCGACGCGC 315
DB 304 CCGGCTGAGATTTTCTTCAAGAAATCTAGGCTTTGAAGATGGCGCATTTGTGAATT 363
QY 316 CACCAAGACACAGCCGAGAGGCGCACTGCGATTTACAGGTGAAGGTGACGAGCACC 375
DB 364 CGTTCAAGATATAGTTTAAGAAATGATTAAGTTTCACTTCAAAAGTGAATAGAGGCAAC 423
QY 376 AACTTCCCGCGGAGCGCGCGTGTATGAAGAACAGAGCGGCGGTGAGAGCCGAGCACC 435
DB 424 GGTTCCTTGTAAACGAGACCGGTGATGCAAAAACCACTCTCGGCAATGAGCCATCGTTT 483
QY 436 GAGGTGTGTACCCCGAAGACGCGTGTCTGTGCGGC 471
DB 484 GAGGTGTGTACATGAACGCGCGTCTGTGTGGC 519
```

```
RESULT 8
US-09-609-161B-31
; Sequence 31, Application US/09609161B
; Patent No. 643682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/277,716A
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(720)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (Insert B)
US-09-277-716-31

Query Match      17.7%; Score 121.6; DB 3; Length 1279;
Best Local Similarity 54.2%; Pred. No. 1.7e-16;
Matches 247; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 16 CTGAAGAGAGATATCCGATCATGATGTATCATGAGGCGACCGTGAAGCGCCACTTCTTC 75
DB 37 CTGAAGAGATATATGTCGCGCAAAAGCTAGCGTTGAAGAAATCGTGAACATCACTTTT 96
QY 76 AAGTGCAGAGGCGAGGCGGAGCGCAACCCCTTCGCCGCGCACCCGAGATGAAATTCAC 135
DB 97 TCCATGGAAGATTTGTGAAAAGGCAATGTAATTTTGAAGAACCAATGATGCAATCCG 156
QY 136 GTGACCGAGGCGCCCCCTTCGCTTCGCTTCGATCTTGAGCCCTTCGCGATGAC 195
DB 157 GTTACAAAGGAGAGTCCGTTGCGCATTCGCTTTCGACATTTTTCATAGCTTTCGAATAC 216
QY 196 GGCAGAGAGACCTTGTGCAACAACCGCGGAGATCCCGACTTCTTCAAGCAGACCTTC 255
DB 217 GGGAAATCGACTTTCACAAATACCGAGACGACATTCGCGACTTTCGATCAATTCATT 276
QY 256 CCCGAGGCTTCACTGCGAGAGAACCACTTACGAGAGCGGCGGATCTTCGACGCGC 315
DB 277 CCGGCTGAGATTTTCTTCAAGAAATCTAGGCTTTGAAGATGGCGCATTTGTGAATT 336
QY 316 CACCAAGACACAGCCGAGAGGCGCACTGCGATTTACAGGTGAAGGTGACGAGCACC 375
DB 337 CGTTCAAGATATAGTTTAAGAAATGATTAAGTTTCACTTCAAAAGTGAATAGAGGCAAC 396
QY 376 AACTTCCCGCGGAGCGCGCGTGTATGAAGAACAGAGCGGCGGTGAGAGCCGAGCACC 435
DB 397 GGTTCCTTGTAAACGAGACCGGTGATGCAAAAACCACTCTCGGCAATGAGCCATCGTTT 456
QY 436 GAGGTGTGTACCCCGAAGACGCGTGTCTGTGCGGC 471
DB 457 GAGGTGTGTACATGAACGCGGCGTCTGTGTGGC 492
```

```

/ TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
/ TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
/ FILE REFERENCE: 24729-121B
/ CURRENT APPLICATION NUMBER: US/09/609,161B
/ CURRENT FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: 09/277,716
/ PRIOR FILING DATE: 1999-03-26
/ PRIOR APPLICATION NUMBER: 60/102,939
/ PRIOR FILING DATE: 1998-10-01
/ PRIOR APPLICATION NUMBER: 60/089,367
/ PRIOR FILING DATE: 1998-06-15
/ PRIOR APPLICATION NUMBER: 60/079,624
/ PRIOR FILING DATE: 1998-03-27
/ NUMBER OF SEQ ID NOS: 32
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 31
/ LENGTH: 1279
/ TYPE: DNA
/ ORGANISM: Ptiliosarcus gurneyi
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (7)..(1720)
/ OTHER INFORMATION: Ptiliosarcus Green Fluorescent Protein (GFP) (Insert B)
US-09-609-161B-31

```

Query Match 17.7%; Score 121.6; DB 4; Length 1279;

Best Local Similarity 54.2%; Pred. No. 1,7e-16; Matches 247; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

```

QY CTGAGAGAGATATGCGCATCAAGATATCATGAGAGGACCGGTGAACGGCCACTACTTC 75
DB 37 CTGAAAGAGATATATGCGCAAAAGCTAGCGTTGAAGAAATCGTGAACAATCAAGTTT 96
QY 76 AAGTGGAGGCGCAGGCGCAGCGCAACCCCTTCCGCGGACCCAGAGCATGAAATCCAC 135
DB 97 TCCATGAAAGATTTTGAAGAAAGGCAATGATATTTTGAAGAAATGATGCAAAATCCG 156
QY 136 GTGACGAGGAGGCGCCCGCTTGCCTTGCCTTGCACATCTTGCGCCCTGCTCGAGTAC 195
DB 157 GTTACAAAGGAGAGTCCGTTGCCATTCGCTTGCACATTTTCCATAGCTTCCATAC 216
QY 196 GGCAGAGAGACCTTCTGTGACCAACCGCGAGATCCCGCATCTTCAAGCAGAGCTTC 255
DB 217 GGAATCGACCTTTCAGAAATACCAAGAGCATTTGCCGACTATTGTTCAATATT 276
QY 256 CCGGAGGCTTCACTCTGGAGAGAACCAACCACTAGAGAGCGCGCATCTGACCGCC 315
DB 277 CCGGCTGATTTTCTTACGAAAGAAATCTACGCTTGAAGATGGCGCATTTGTGACAT 336
QY 316 CACGAGAGACACGCGTGGAGGCACTGCTGATCTACAAGGTGAAGTGACAGGACCC 375
DB 337 CGTTCAAGATATTAAGTTTGAAGATGATTAAGTTCCACTCAAAAGTGAAGTGAAGGCAAC 396
QY 376 AACTTCCCGCGCGCGCGCGCGCGATGAAAGAACAGAGCGCGCTGAGCCAGCACCC 435
DB 397 GGTTCCTTAGTAAACGAGACCGGATGATGACAAAAGCCATCTCGGCATGAGGCAATGCTTT 456
QY 436 GAGGTGTGTATCCCGAGAAACGCGCTGCTGCGGC 471
DB 457 GAGGTGTGTATCAATGAACAGCGCGCTTGTGTTGGGC 492

```

RESULT 9

```

US-09-172-063-14
/ Sequence 14, Application US/09172063
/ Patent No. 6150176
/ GENERAL INFORMATION:
/ APPLICANT: Tsielen, Roger Y.
/ APPLICANT: Miyawaki, Atsushi
/ APPLICANT: Llopis, Juan
/ APPLICANT: Wachter, Rebekka M.
/ APPLICANT: Remington, S. James

```

```

/ TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
/ TITLE OF INVENTION: MEASURING THE PH OF A BIOLOGICAL SAMPLE
/ FILE REFERENCE: 07257/071001
/ CURRENT APPLICATION NUMBER: US/09/172,063
/ CURRENT FILING DATE: 1998-10-13
/ EARLIER APPLICATION NUMBER: 09/094,359
/ EARLIER FILING DATE: 1998-06-09
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 14
/ LENGTH: 720
/ TYPE: DNA
/ ORGANISM: Aequorea victoria
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (0)..(0)
/ OTHER INFORMATION: EYFP-V68L/Q69K
US-09-172-063-14

```

Query Match 16.1%; Score 110.6; DB 3; Length 720;

Best Local Similarity 55.2%; Pred. No. 2.8e-14; Matches 262; Conservative 0; Mismatches 204; Indels 9; Gaps 2;

```

QY 13 CTGCTGAGAGAGATATGCGCATCAAGATATCATGAGAGGCAACCGTGAACGGCCACTAC 72
DB 22 CTGTTCAACCGGAGTGTGCTCCATCTGCTGAGCTGAGACGGAGCGTAAACGGCCCAAG 81
QY 73 TTCAAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
DB 82 TTCAAGCTGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 141
QY 133 CAGTGACCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 192
DB 142 ATCTGACACAC--GGGAGAGCTGCGCCGCTGAGCCACCTGTGTGACACCTTCCGCG 198
QY 193 TACGAGCAGAGACCTTCTGTGACCAACCGCGAGAT-----CCCGCATTTTCAAG 246
DB 199 TACGCGCTGAGATGCTTCCCGCTTACCCCGTACCCGACCATGAGACAGACGACATCTTCAAG 258
QY 247 CAGAGCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 306
DB 259 TCCGCGATGCGGAGAGGAGTACGTCAAGAGCGCACCATTTTCAAGAGCAGAGGCAAC 318
QY 307 CTGACCGCCACACAGAGACACAGCTGAGGAGCACTGCTGATCTACAAGGTGAAGTGTG 366
DB 319 TACAAGACCGGCGGAGGTGAAGTTCAAGGCGACACCTGTGTAACCGCATGACGTG 378
QY 367 CAGGACCACTTCCCGCGAGCGCGCGCGCTGATGAAGAACAGAGCGCGCTGGAGAG 426
DB 379 AAGGCGATGACTTCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 438
QY 427 CCCAGACCGAGAGTGTATACCCCGAGAACGCGCGTGTGGCGCGGAGACGTGA 481
DB 439 AACAGCCACACGTATATATATGAGCGGACAGAGAAAGGAGGATCAAGGTGA 493

```

RESULT 10

```

US-09-316-919-15
/ Sequence 15, Application US/0916919
/ Patent No. 6469154
/ GENERAL INFORMATION:
/ APPLICANT: Tsielen, Roger Y.
/ APPLICANT: Baird, Geoffrey
/ TITLE OF INVENTION: FLUORESCENT PROTEIN INDICATORS
/ FILE REFERENCE: 07257/073001
/ CURRENT APPLICATION NUMBER: US/09/316,919
/ CURRENT FILING DATE: 1999-05-21
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 720
/ TYPE: DNA
/ ORGANISM: Aequorea victoria

```

US-09-316-919-15

Query Match 16.1%; Score 110.6; DB 4; Length 720;
 Best Local Similarity 55.2%; Pred. No. 2.8e-14;
 Matches 262; Conservative 0; Mismatches 204; Indels 9; Gaps 2;

QY 13 CTGCTGAAGAGAGATATGCCATCAAGATGATCATGAGAGGCAACCGTGAACGGCACTAC 72
 DB 22 CTGTTACCGGGGAGTGCCCATCTGTGTGAGCTGACGCGCAAGTAAACGGCCACAAG 81
 QY 73 TTCAAGTCGAGGGGCGAGGGCGCAACCCCTTTCGCGGCAACCAAGACATGAGAAATC 132
 DB 82 TTCAAGCTGTCCGGCGAGGGCGAGGGCGATGCACTACGCAAGCTGACCTGAAAGTTC 141
 QY 133 CAGGTGACGAGGGCGCCCGCTGCGCTTTCGACATCTGTGCCCCCTGTGCGAG 192
 DB 142 ATCTGACACAC---CGCAAGCTGGCCCTGTGGCCACCTTCGTGACCACTTTCGAGC 198
 QY 193 TACGGCAGCAGGACCTTTCGTGACCAACACCGCCGAGAT-----CCCGACTTTCAG 246
 DB 199 TACGGCTGAAGTGCTTCCCGCTACCCCAACACATGAAGACACGACTTCTTCAAG 258
 QY 247 CAGAGCTTCCCGAGGGGCTTTCACCTGGAGAGAAACCAACCTACGAGAGACGGCGATC 306
 DB 259 TCGGCATGCCCGAGAGGCTACCTCCAGAGAGCACCATCTTTCAGAGACGAGCGCAC 318
 QY 307 CTGACCGCCCAACGAGACACGACCTGAGAGGCAACGCTGATCTACAGAGTGAAGGTG 366
 DB 319 TACAGACCCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGATCGAGCTG 378
 QY 367 CACGGCACCACCTTCCCGCGAGCGCCCGTATGAAGAAAGAGGGCGGCTGGAG 426
 DB 379 AAGGGCATGCACTTCAAGAGAGACGGCAACATCTGGGCAACACTGATTAACACTAC 438
 QY 427 CCCGACCCGAGGTGTGTATCCCGAAGACGGGTCTCTGTGGCCGGAACGTGA 481
 DB 439 AACAGCCACAAGCTATATATATGCGGACGACAGAAAGACGGCATCAAGTGA 493

RESULT 11

US-09-602-641-14
 ; Sequence 14; Application US/09602641
 ; Patent No. 6608189
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Miyawaki, Atsushi
 ; APPLICANT: Llopis, Juan
 ; APPLICANT: Wachter, Rebekka M.
 ; APPLICANT: Remington, S. James
 ; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
 ; FILE REFERENCE: 07257/071001
 ; CURRENT APPLICATION NUMBER: US/09/602,641
 ; PRIOR FILING DATE: 2000-06-22
 ; PRIOR APPLICATION NUMBER: 09/172,063
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 720
 ; TYPE: DNA
 ; ORGANISM: Aequorea victoria
 ; NAME/KEY: misc.feature
 ; FEATURE:
 ; LOCATION: (0) ... (0)
 ; OTHER INFORMATION: EYFP-V68L/Q69K
 US-09-602-641-14

Query Match 16.1%; Score 110.6; DB 4; Length 720;
 Best Local Similarity 55.2%; Pred. No. 2.8e-14;
 Matches 262; Conservative 0; Mismatches 204; Indels 9; Gaps 2;
 QY 13 CTGCTGAAGAGAGATATGCCATCAAGATGATCATGAGAGGCAACCGTGAACGGCACTAC 72

DB 22 CTGTTACCGGGGAGTGTCCTCATCTGTGTGAGACTGAACGGGACGTAAGCGCCACAG 81
 QY 73 TTCAAGTGGAGAGGGGAGAGGAGACGACACCCCTTTCGCGGCAACCAAGAGATGAAATC 132
 DB 82 TTCAAGCTGTCCGGCGAGGGCGAGGGCGATGCACTACGCAAGCTGACCTGAAAGTTC 141
 QY 133 CAGGTGACGAGGGGCGCCCGCTGCGCTTTCGACATCTGTGCCCCCTGTGCGAG 192
 DB 142 ATCTGACACAC---CGCAAGCTGGCCCTGTGGCCACCTTCGTGACCACTTTCGAGC 198
 QY 193 TACGGCAGCAGGACCTTTCGTGACCAACACCGCCGAGAT-----CCCGACTTTCAG 246
 DB 199 TACGGCTGAAGTGCTTCCCGCTACCCCAACACATGAAGACACGACTTCTTCAAG 258
 QY 247 CAGAGCTTCCCGAGGGGCTTTCACCTGAGAGAAACCAACCTACGAGACGGCGCATC 306
 DB 259 TCGGCATGCCCGAGAGGCTACCTCCAGAGGCGCACCATCTTTCAGAGACGAGCGCAC 318
 QY 307 CTGACCGCCCAACGAGACACAGCTTGAAGGCACTGCTGATCTACAGAGTGAAGGTG 366
 DB 319 TACAGACCCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGATCGAGCTG 378
 QY 367 CACGGCACCACCTTCCCGCGAGCGCCCGTATGAAGAAAGAGCGGCGGCTGGAG 426
 DB 379 AAGGGCATGCACTTCAAGAGAGACGGCAACATCTGGGGGCAAGAGTGAAGTCAACTAC 438
 QY 427 CCCGACCCGAGGTGTGTATCCCGAAGACGGGTCTGTGGCCGGAACGTGA 481
 DB 439 AACAGCCACAAGCTATATATATGCGGACGACAGAAAGACGGCATCAAGTGA 493

RESULT 12

US-09-459-956-2
 ; Sequence 2; Application US/09459956
 ; Patent No. 6342379
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Gonzalez, Ili, Jesus E.
 ; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
 ; FILE REFERENCE: OPTICAL METHODS
 ; CURRENT APPLICATION NUMBER: US/09/459,956
 ; PRIOR FILING DATE: 1999-12-13
 ; PRIOR APPLICATION NUMBER: 08/765,860
 ; PRIOR FILING DATE: 1999-05-08
 ; PRIOR APPLICATION NUMBER: 08/481,977
 ; PRIOR FILING DATE: 1995-06-07
 ; PRIOR APPLICATION NUMBER: PCT/US96/09652
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 690
 ; TYPE: DNA
 ; ORGANISM: Anemonia majano
 US-09-459-956-2

Query Match 16.0%; Score 109.8; DB 4; Length 690;
 Best Local Similarity 57.0%; Pred. No. 4e-14;
 Matches 223; Conservative 0; Mismatches 162; Indels 6; Gaps 1;
 QY 44 ACATGAGAGGAGCGGTGAGGCACTTCAATGTGAGAGGGGAGAGCGGCAAC 103
 DB 50 ATATGAGTGTGTGTCAATGGCATTTACTTACCGTAAAGGTGAAGGAGGCAAGCGGAGAC 109
 QY 104 CCTGCGCGGACCCCAAGCATAG-----AATCACTGACCGAGGGCGCCCTGTC 157
 DB 110 CATAGAAAGGAGCGGACCTTCACTTTAAAGTACACCATAGGCCAAGCGGTGGCCCTTG 169
 QY 158 CTTTGCCTTGCATCTGCGCCCGCTGTGAGTACGCGACAGACGACCTTGTGACAC 217
 DB 170 CATCTCCTTGGACATATATCTACAGTGTCAATATATGAAATGATGATCTTACTGGCT 229

QY 218 AACCCGCGGATCCCGACTTCTTCAAGCAGAGCTTCCCGAGGCGTTCACTGGAGAG 217
DB 230 ATCTACACGATATGCCGACTATTTCAACACAGCATTTCTGACGGAAATGTATGAAA 289
QY 278 GAACCAACCACTTCAAGAGAGCGGCGCATCTGACCGCCCAACGAGACACAGCTTGAGG 337
DB 290 GGACTTTTACCTTATGAGATGAGGAGATTGCTACAGCCAGTTGGAAATTAAGCTTAAAG 349
QY 338 GCACTGCTGATCTTCAAGGTGAGTGACGCGACCACTTCCCGCGCGACGCGCCG 397
DB 350 GCACTGCTTTGAGCACAATCCAGTTTATGAGTGAATTTCTGCTGATGAGCTTG 409
QY 398 TGATGAAGAACAGAGCGCGCGCTGGAGCC 428
DB 410 TGATGGCAAGAGACAACTGTTGGGACC 440

RESULT 13
US-09-513-783A-43
; Sequence 43, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliani, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(717)
; OTHER INFORMATION: Description of Artificial Sequence: EYFP
US-09-513-783A-43

Query Match 15.9%; Score 109; DB 4; Length 717;
Best Local Similarity 54.9%; Pred. No. 5.9e-14;
Matches 261; Conservative 0; Mismatches 205; Indels 9; Gaps 2;

QY 13 CTGCTAAGGAGATGCGCATCAAGTATCATGAGGCGACCGTGAACGGCCACTAC 72
DB 22 CTGTTACCGGGGTGGTGGCCATCTGTGAGCTGACGCGACGTAAACGGCCACAAG 81
QY 73 TTCAAGTGCAGAGGCGAGGCGAGCAACCCCTTGC CGGACCCAGAGCATGAGATC 132
DB 82 TTCAAGGTGTCCGCGAGGCGAGGCGAGTCCACCTACGCAAGCTGACCTGAAAGTTT 141
QY 133 CACGTACCGAGGCGCGCCCTTGCCTTGCATCTGTGCGCCCTGCTGCGAG 192
DB 142 ATCTGACACAC---CGGCAAGCTGCCGCTGCGCCCACTCTGTGACCACTTCCGCG 198
QY 193 TACGGAGACGACCTTCTGTGACCAACCGCCGAGAT-----CCCGACTTTCTTCAAG 246
DB 199 TACGGCTGTGAGTGTCCGCGCTTACCCGACCACTGAGAGACAGCACTTCTTCAAG 258
QY 247 CAGAGCTTCCCGAGGCGCTTCACTGAGAGAGAACCACTACGAGAGAGCGGCGATC 306
DB 259 TCCGCCATGCGCGAAGGCTACGTCAGAGAGCGACCATCTTCTTCAAGAGACGCGCAAC 318
QY 307 CTGACCGCCACAGAGACACAGCTGAGAGGCAACTGCTGATCTTCAAGGTGAAGGTG 366
DB 319 TACAAAGCCCGCGAGGTGAAGTTCAGAGGCGACACCTGTGTAACCGCATGAGCTG 378
QY 367 CACGGACCAACTTCCCGCGCGAGCGCCCTGTATGAAGAACAGAGCGGCGCTGGAG 426
DB 379 AAGGGATGACTTCAAGAGAGCGGCAACATCTGGGGGCAAGGCTGAGTGAACACTAC 438

QY 427 CCCAGACCGAGGTGTATACCCCGAAGCGCGTGTCTGCGCGCGGAAACGTGA 481
DB 439 AACAGCCACAAGCTTATATCATGCGCGACAAAGAGAAAGCGGCATCAAGGTGA 493

RESULT 14
US-08-911-825-3
; Sequence 3, Application US/08911825
; Patent No. 6054321
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,825
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..720
US-08-911-825-3

Query Match 15.9%; Score 109; DB 3; Length 720;
Best Local Similarity 54.9%; Pred. No. 5.9e-14;
Matches 261; Conservative 0; Mismatches 205; Indels 9; Gaps 2;

QY 13 CTGCTAAGGAGATGCGCATCAAGTATCATGAGGCGACCGTGAACGGCCACTAC 72
DB 22 CTGTTACCGGGGTGGTGGCCATCTGTGAGCTGAGCGGACGTAAACGGCCACAAG 81
QY 73 TTCAAGTGCAGAGGCGAGGCGAGCAACCCCTTGC CGGACCCAGAGCATGAGATC 132
DB 82 TTCAAGGTGTCCGCGAGGCGAGGCGAGTCCACCTACGCAAGCTGACCTGAAAGTTT 141
QY 133 CACGTACCGAGGCGCGCCCTTGCCTTGCATCTGTGCGCCCTGCTGCGAG 192
DB 142 ATCTGACACAC---CGGCAAGCTGCCGCTGCGCCCACTCTGTGACCACTTCCGCG 198
QY 193 TACGGAGACGACCTTCTGTGACCAACCGCCGAGAT-----CCCGACTTTCTTCAAG 246
DB 199 TACGGCTGTGAGTGTCCGCGCTTACCCGACCACTGAGAGACAGCACTTCTTCAAG 258
QY 247 CAGAGCTTCCCGAGGCGCTTCACTGAGAGAGAACCACTACGAGAGAGCGGCGATC 306
DB 259 TCCGCCATGCGCGAAGGCTACGTCAGAGAGCGACCATCTTCTTCAAGAGACGCGCAAC 318
QY 307 CTGACCGCCACAGAGACACAGCTGAGAGGCAACTGCTGATCTTCAAGGTGAAGGTG 366

Db 319 TACAGACCCGCGCGAGGTGAAGTTGAGGGGACACCTGTGTAAACCGCATGAGCTG 378
 QY 367 CACGGACCAACTTCCCGCGACGCGCCCGTGAAGAAACAAGAGCGCGCTGGAG 426
 Db 379 AAGGCATCGACTTCAAGAGACGAGCAATCTGGGCGACAGCTGAGTCAACTAC 438
 QY 427 CCCAGCACCGAGGTGTACCCGAGAACGCGGTGTGTGGCGCGGAACGTGA 481
 Db 439 AACAGCCAAAGTCTATATCATGTGCGGACAGACAGAAAGCGCATCAAGTGA 493

RESULT 15

US-08-974-737-3
 ; Sequence 3, Application US/08974737
 ; Patent No. 6077707

GENERAL INFORMATION:

; APPLICANT: Tsjen et al., Roger Y. et al.
 ; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
 ; NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Suite 1400
 ; CITY: La Jolla
 ; STATE: CA

; COUNTRY: USA
 ; ZIP: 92037

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/974,737
 ; FILING DATE:

CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/911,825
 ; FILING DATE: 15-AUG-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Lisa A.
 ; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 07257/056001
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619/678-5070
 ; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 720 base pairs
 ; TYPE: nucleic acid

; STRANDEDNESS: single
 ; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS
 ; LOCATION: 1..720

US-08-974-737-3

Query Match 15.9%; Score 109; DB 3; Length 720;
 Best Local Similarity 54.9%; Pred. No. 5.9e-14;

Matches 261; Conservative 0; Mismatches 205; Indels 9; Gaps 2;

QY 13 CTGCTGAAGAGAGTATGCGCATCAAGATGTATCATGAGGGGACCGTGAACGGGCACTAC 72
 Db 22 CTGTTACACCGGGGGTGGTCCCATCTGTGTGAGCTGAGCGGCAAGCTAAACGGCCACAG 81
 QY 73 TTCAAGTCCGAGGGGCGGAGCGGCAACCCCTTGCCTGCGGACCCGAGCATGAGATTC 132
 Db 82 TTCAAGTCTCGGCGGCGGCGGAGTGCACCTTACGGGCAAGCTGACCTTGAAGTTC 141
 QY 133 CACGTGACCGAGGGCGGCGGCTTGCCTTGCATCTGAGCATCTGAGCCCTGCTGCGAG 192
 Db 142 ATCTGACACAC---CGGCAAGCTGCGGCTGCGCCACCTCTGAGCACACCTTGGGC 198

QY 193 TACGGCAGAGACCTTGTGTGACACACCGCGAGAT-----CCCGACTTCTTCAAG 246
 Db 199 TACGGGTGCACTGTCTTCCCGCTACCCGACACATTAAGACACAGACTTCTTCAAG 258
 QY 247 CAGAGCTTCCCGAGGGTTTCACTGTGAGAGAAACCACTTACGAGAGCGGCGATC 306
 Db 259 TCGGCATGTCCCGAAGGTACTGTCTCAGAGCGGCAATCTTGAAGGAGACGCGCAC 318
 QY 307 CTGACCGGCCACAGACACCAAGCTGAGGGCAACTGTGTATCTAACAGTGAAGTGTG 366
 Db 319 TACAAAGACCCGCGCGAGTGAAGTTGAGGGGACACCTGTGTAAACCGCATGAGCTGT 378
 QY 367 CACGGACCAACTTCCCGCGACGCGCCCGTGAAGAAACAAGAGCGCGCTGGAG 426
 Db 379 AAGGCATCGACTTCAAGAGACGAGCAATCTGGGCGACAGCTGAGTCAACTAC 438
 QY 427 CCCAGCACCGAGGTGTACCCGAGAACGCGGTGTGTGGCGCGGAACGTGA 481
 Db 439 AACAGCCAAAGTCTATATCATGTGCGGACAGACAGAAAGCGCATCAAGTGA 493

Search completed: July 15, 2004, 14:06:51
 Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2004, 11:15:43 : Search time 2886 Seconds
(without alignments)
7108.565 Million cell updates/sec

Title: US-09-976-673-11

Sequence: 1 atcgctgagcgcctgctgaa.....tgcccgagaaggcgaactga 687

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gse_hum:*
18: em_gse_inu:*
19: em_gse_pln:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_man:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rtd:*
26: em_gse_pbg:*
27: em_gse_vrl:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.4	14.5	550	13	BQ551001 H4006G01-
2	71	10.3	645	9	AI389106 GH20192.5
3	68.6	10.0	869	14	CK159167 FGAS04056
4	67.6	9.8	639	9	AJ615477 AJ615477

5	67.4	9.8	675	12	B1382638 BFLG2_000
6	66.8	9.7	944	29	CC649516 OGCLU16TH
7	66.4	9.7	700	14	CF864358 psz25009xP
8	65.8	9.6	624	24	B2422085 id49d07_b
9	65.8	9.6	982	28	CC424389 PUHKK53TR
10	65.2	9.5	794	14	CB651092 OSJNEB150
11	64.8	9.4	697	14	B0626888 UI-H-FTO-
12	64.6	9.4	821	14	CK199961 FGAS00846
13	64.2	9.3	613	12	B1386976 BFL26_001
14	64.2	9.3	643	12	B1386999 BFL26_001
15	64	9.3	681	29	CNS0280D Tetradon
16	64	9.3	688	29	CG038305 PUENQ30TD
17	63.8	9.3	411	10	BF201975 WHB1759-1
18	63.8	9.3	600	10	BE417750 MUG024..B0
19	63.8	9.3	772	13	BQ752733 WHE4118_E
20	63.6	9.3	872	28	CC413363 PUHPU48TD
21	63.6	9.3	956	29	CG365299 OGFWF62TV
22	63.4	9.2	925	29	CNS0081P Drosoph11
23	62.8	9.1	1006	29	CC719047 OGUN81TV
24	62.6	9.1	599	29	A1832082 tcl2c04..x
25	62.6	9.1	689	12	B1380055 BFLG1_001
26	62.4	9.1	1036	14	CK162903 FGAS01551
27	62.2	9.1	638	12	B3269933 B3269933
28	62.2	9.1	742	14	CB649099 OSJNEB12N
29	62.2	9.1	780	14	CF554490 G9W4_Inf
30	62	9.0	786	14	CB646357 OSJNEB08M
31	62	9.0	982	13	BK415111 BX415111
32	61.8	9.0	557	12	B3233355 B3233355
33	61.6	9.0	590	13	CA030137 HX06D04T
34	61.4	8.9	540	12	BM135283 WHE0498_B
35	61.4	8.9	634	14	CA254345 SCBFL411
36	61.2	8.9	724	14	CF871694 tlc028xP
37	61.2	8.9	739	14	CB645917 OSJNEB07P
38	61.2	8.9	777	14	CB901927 tlc028xP
39	61	8.9	571	12	B1387406 BFL26_002
40	61	8.9	661	12	B1385436 BFL26_000
41	61	8.9	687	12	B1387917 BFL26_002
42	61	8.9	813	14	CB657967 OSJNEC13K
43	60.8	8.9	502	12	B1354401 6HRm115_6
44	60.6	8.8	559	14	CF847678 p6MA002x8
45	60.6	8.8	759	28	B2577913 msh2_5633

ALIGNMENTS

RESULT 1
LOCUS BQ551001 550 bp mRNA linear EST 20-JUN-2002
DEFINITION H4006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
H4006G01 5', mRNA sequence.
ACCESSION BQ551001
VERSION BQ551001.1 GI:21451887
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 550)
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
Martin,P.R., Stagg,C.A., Baesey,U., Abba,K., Hamatani,T.,
Kargul,G.J., Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse
CDNA clone set
JOURNAL Genome Res. 12 (12), 1999-2003 (2002)
MEDLINE 22354164
PUBMED 12466305
COMMENT Other ESTs: H4006G01-3
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please visit http://lgsun.gic.nia.nih.gov/cDNA/NIA_7_4k.html for details.
 Place: H4006 row: 5 column: 01
 Seq primer: -21M13 Reverse
 High quality sequence stop: 550
 POLYA=No.

FEATURES

Location/Qualifiers
 1. 550
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /db_xref="taxon:10090"
 /clone="H4006G01"
 /sex="mixed"
 /dev_stage="mixed"
 /lab_host="DH10B"
 /clone_id="NIA Mouse 7.4K cDNA Clone Set"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 7,407 clones from more than 20 cDNA libraries."

ORIGIN

Query Match 14.5%; Score 99.4; DB 13; Length 550;
 Best Local Similarity 55.3%; Pred. NC. 8.6e-05;
 Matches 215; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

13 CTGCTGAAGAGAGATGCGCATCAAGTGTACATGAGAGGCACTGTGAAGGCCACTAC 72
 Db CTGTTACCGGGGTGTGCTTCATCTGTGACAGTGTGACGAGTAAAGGCCACAG 218
 Oy TTCAAGTGAAG 132
 Db TTCAAGTGTGCGGAG 278
 Oy 133 CAGCTGACCGAAGGCG--GCCCTCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 189
 Db 279 ATCTGACACCGAG 338
 Oy 190 GAGTACGAG 249
 Db 339 GCGGTGACAGTGTGCTTGTGAGCGCTACCCGACCAAGTGTGAGAGAGAGAGAG 398
 Oy 250 AGCTTCCCGAG 309
 Db 399 GCGATGCCCGAG 458
 Oy 310 ACCGCCACAG 369
 Db 459 AAGATCCCGCGAG 518
 Oy 370 GGCACCAACTTCTCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
 Db 519 GGCATGACATTCAG 547

RESULT 2
 A1389106 645 bp mRNA linear EST 23-APR-2001
 LOCUS GH20192.5prlme GH Drosophila melanogaster head pot2 Drosophila
 DEFINITION A1389106
 ACCESSION A1389106
 VERSION A1389106.2 GI:13758704
 KEYWORDS EST
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 645)
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 BDGP/HMMI Drosophila Est Project

JOURNAL

COMMENT
 Unpublished (2001)
 On Jan 28, 1999 this sequence version replaced gi:4203117.
 Other ESTs: GH20192.3prlme
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: GH.201 row: H column: 8
 High quality sequence stop: 627
 POLYA=No.

FEATURES

Location/Qualifiers
 1. 645
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="GH20192"
 /sex="male and female"
 /dev_stage="adult"
 /lab_host="DH5 - alpha"
 /clone_id="GH Drosophila melanogaster head pot2"
 /note="Organ: head; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."

ORIGIN

Query Match 10.3%; Score 71; DB 9; Length 645;
 Best Local Similarity 48.6%; Pred. NC. 0.99;
 Matches 194; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

Oy 62 AGCGCACTTCAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
 Db 200 AGCACTCTCAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 259
 Oy 122 GATGAGATTCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
 Db 260 ACAACCAAGTGTTCCTCACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 319
 Oy 182 CTTGCTGAGTACGAG 241
 Db 320 ACCACCAACTACGAG 379
 Oy 242 TCAAGCAGAGCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
 Db 380 CCGACCAACAG 439
 Oy 302 GCATCTGAG 361
 Db 440 ACAACAG 499
 Oy 362 AGGTGACGAG 421
 Db 500 ACCACCAACCAACTACTACTACTACTACTACTACTACTACTACTACTACTACT 559
 Oy 422 GGGAGCCAG 460
 Db 560 ACAACCAACAG 598

RESULT 3
 CK159167/c 869 bp mRNA linear EST 05-DEC-2003
 LOCUS FGAS040564 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,
 DEFINITION CK159167
 ACCESSION CK159167
 VERSION CK159167.1 GI:38985053
 KEYWORDS EST
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticaceae; Triticum.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 869)	Allard, F., Crosby W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genselin, B., Graf, R., Gulick, P., Hyman, L.D., Lacroche, A., Link, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D., Pennikel, C., Roach, J. L. and Sahhan, F.	Functional Genomics of Abiotic Stress in Wheat and Canola Crops		Unpublished (2003)
	Contact: Mm L Crosby			

Email: fgas.estr@cs.usnrc.ca
This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [128,636].
Plate: Talt537 row: N column: 23.

FEATURES	Location/Qualifiers
source	1. .869

```

/organism="Triticum aestivum"
/mol_type="rRNA"
/cultivar="wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS; Talt5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression
subtractive hybridization) cDNA library from genotype
PI178383 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction
heated) and subtracted against genotype Norstar cold
hardened at 2 C for 1 day (24 h) (driver). Modified Smart
cDNA (Clontech) priming and non-directional cloning"

```

ORIGIN

Query Match	10.0%;	Score 68.6;	DB 14;	Length 869;
Best Local Similarity	44.2%;	Pred. No. 2.3;		
Matches 284;	Conservative	0;	Mismatches 359;	Indels 0;
				Gaps 0;

QY 24 GAGTAAAGCCATGAAGATGTACATG9AGG6ACACCGGAAAGGCACTTCTTCAATGGGA 83
Db 807 GAGGATCAACCAACAGAGCAAGAAACAACAACAACAACAACAAGAGCAAGCAACACGGA 744
QY 84 GGGGAGAGGCGAAGCGCAACCCCTTCCCGCGCAACCAAGACATGAAATTCACAGTGAACGA 144

QY 144 GGGGGCCCCCTGCTTGGCAATCTGGGCCCCCTGTGAGATAGCGAGAG 207
D6 687 CAAACACACACGACACACACACACACACACACACACACACACACACACAC 622

Oy 204 GACCTTCGTGCAACACCCGGAGATCCCCGATTCTTCAAGCAGAGCTTCCCAGGG 26
| | | | | | | | | | | | | | |
Db 627 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 56

Oy 264 CTTACCTGGAGAGAAACCACCACTTAGAGAGCGGCGCATCTTGACGCCCCACAGGA 32
|||
Db 567 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 50

Oy 324 CACCAAGCTGGAGGGCAATGCTGATCTACAAGGTGAAGTGACGGCACTTCCC 38

Db 507 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 441

Oy 384 CGCCACGGCCCCGTGATGAAGAACAAGAGCGGCGCTGGAGGCCAGACCCGAGGTGTT 44
 Db 447 CAACACGACGACGACGACCAACGACGACGACGACGACGACCAACACGACGACCAACGACG 38

Oy 444 GTACCCCGAGAA CGGCGTGTCTGTGGCGCGGAACGTATGGCCCTGAAGSTGGCGACCG 50
|||
Db 387 CAACAACAACGACAA CAAACAACAACAACAACGACCAACGACAA GACAA 322

Qy	504	GGACCTGATGTCGCACACTAACAGGTACCGAGGAAGGCCGCGCCCTTGAC	563
Db	327	CACCAACCAACCAACACACACGACACACACACACACACACACACACACAC	266
Qy	564	CATGCCCGGCTTCACCTTCAACCGACATCCGGCTCCAGATGCTGGCGAAGAAAGACGA	622
Db	267	CACCAACCAACCAACACACACACACACACACACACACACACACACACACACGA	208
Qy	624	GTACTTCGAGCTGTACGAGGCCAGCGCTGAGCCCGTACAGCGAC	666
Db	207	CGACACGACGCTCAAGCATACACAGAACACACACACACAC	165

LOCUS	DEFINITION
AJ615477	RESULT 4
AJ615477	LOCUS
AJ615477	639 bp mRNA linear EST 11-DEC-2003
AJ615477	Triticum turgidum subsp. durum etiolated seedling 20 day
Triticum turgidum subsp. durum cDNA clone 11702R, mRNA sequence.	

ACCESSION	AJ615477
VERSION	AJ615477.1
	GI:39732404

KEYWORDS. EST. *Triticum turgidum* subsp. *durum* (durum wheat) *Triticum turgidum* subsp. *durum*

ORGANISM
Triticum turgidum subsp. durum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1 (bases 1 to 639)

REFERENCE

JOURNAL
Unpublished (2003)
Contact: Cifarelli RA

AUTHORS
Cifarelli, R.A., D'Onofrio, O. and Iantia, G.

TITLE
Expressed Sequences Tags (ESTs) Library from totipotent CDNA of durum wheat

**Biotechnology
Metapontum Agrobios
S.S. Jonica 106 Km 448.2, 75010 Metaponto (MT), Italy
Location/Qualifiers**

FEATURES
SOURCE

```

/organism="Triticum turgidum subsp. durum"
/mol_type="mRNA"
/cultivar="Ofanto"
/sub_species="durum"
/db_xref="taxon:4567"
/clone="11702R"
/tissue_type="etiolated seedling"
/dev_stage="20 day"
/clone_lib="Triticum turgidum subsp. durum etiolated
seedling 20 day"

```

Query Match	9.84;	Score 67.6;	DB 9;	Length 639;
Best Local Similarity	52.9%;	Pred. No. 3;		
Matches 145; Conservative	0;	Mismatches 129;	Indels 0;	Gaps 0;

```
Oy      58 GTGACCGGCCTACTTCAAGTCGACGGGCGAGGGCGACGGCAACCCCTTGCGCGGACC 11
        ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     294 GTCTGTGGGCGTCAACAAGAAGTCGCTGTGCGTGTGCTCAAGAACCGCGACGACGCCAAC 353
```

Oy 118 CAGAGCATGAGAAATCCACGTGCAGCCGAGGGCGCCCCCTGCCCTTGCGCTTCGACATCCTG 177
| | | | | | | | | | | | | | | | | | | | | |
Db 354 CTCGGGATCAAGTTCACGCTCACCCGGCCATTGAACCTCCCCTCCGGCTGCAACATCCCG 413

Oy 178 GCCCCTGCTGCAGTACGGCAGCAGGACTTTGTGTCAACCACCGCCGAGATCCCCGAC 233
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 GGCACCTTCTCCGACTGCCCAAGATCCTCAACATTCGCGCGGACTCCAAGAGGCGGAG 473

Oy 238 TTCTTCAAGCAGAGCTTCCCCGAGGGCTTCACTGGGAGAGAAACCAACCTACGAGAAC 297
 |||||
 Db 474 ATCTTCAGACGTAACGGATCGACACGAGGGCAAGAACGCCACCGCCGGCGGACGCGCC 533

QY 258 GGGCGCATCTGACCGGCCACGAGACACCAAGCC 331
| | | | | | | | | | | | | | | | | | | | | |
Db 534 GCCGTACCCGGTACTCCGGCGGGGAAGAGCCGCC 567

Fri Jul 16 08:44:56 2004

us-09-976-673-11.txt

Page 4

RESULT 5
LOCUS BI382638
DEFINITION BI382638 675 bp mRNA linear EST 26-AUG-2003
BFL2.000815 Amphioxus 5-6 hrs cDNA library (Name convention: BFL2
or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498F0258 5',
mRNA sequence.
VERSION BI382638
KEYWORDS BI382638.1 GI:30917704
SOURCE EST.
ORGANISM Branchiostoma floridae (Florida lancelet)
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.

REFERENCE 1 (bases 1 to 675)
AUTHORS Panopoulou,G., Hemmig,S., Groth,D., Krause,A., Pousetka,A.J.,
Herwig,R., Vingron,M. and Lehrach,H.
TITLE New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
Genomes
JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE 22683279
COMMENT 12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhest.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting
(ONPFI) to reduce sequencing redundancy. According to the ONPFI
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCGAGCTTACCTTACCTTACCTTCCGGCTG 3' (M13RSP)
BACKWARD: 5' GCTATTACGACCTGCGAAGGGGATGTG 3' (M13FSP)
Insert length: 1 Std Error: 0.00
Seq primer: 5'-CCGCTCGGATTCCTCGGGT-3' pSPORT3/86
High quality sequence, stop: 675.
Location/Qualifiers
1..675
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGP498F0258"
/class_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, Xli blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFL2 or MPMGP498)"
/note="Vector: pSPORT1; Site 1: SalI, KpnI, SmaI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
(5') directionally cloned in pSPORT1 vector using a NotI
(5')/PAC1/TAATGATCGGACCGGCGCC (T)15-3' and a SalI 5'-
TCGACCAACGCGCTCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 9.8%; Score 67.4; DB 12; Length 675;
Best Local Similarity 48.1%; Pred. No. 3.2;
Matches 191; Conservative 0; Mismatches 206; Indels 0; Gaps 0;
36 CAAGATGATCATGAGGACCGGTGAACGCACTTCAATGTCGAGAGGCGGACGCGCA 95
|||||

Db 106 GCAGATCCACCTTTCACGGCTCCATCCAAAGCCACAGATTTCGACATGCTGGGGGAGGAA 165
Qy 96 CCGCAACCCCTTCCCGGACCCAGAGCATGAGATTCACATGACCGAGGCGCCCTT 155
Db 166 AGGGAGACCCGAAGCGCGCTGCTGGTGACACAGGAAATCCACCAAGGGTCCCTGAA 225
Qy 156 GCCCTTGGCTTTCGACATCTTGGCCCTTCTGCGATGACGAGCAGACGACTTCTGCA 215
Db 226 GTTCTCCCTTCTTGAATGATCCCTCCACTCGGGTACGGGTACTACGATCTCTCCCTA 285
Qy 216 CCAACACCGCGAGATCCCGGACCTTCTTCAAGAGAGCTTCCCGAGGGCTTCACTTGGGA 275
Db 286 CCGGACGAGACCGCTGCTTTTCAAGTCTTCATGTTGAAGATGCGGGATGCGACTTA 345
Qy 276 GAGAACACCAACCTTACGAGAGCGCGCATCTTACCGCCCAACGAGACACGAGCTTGA 335
Db 346 CCGCGTTCATCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 405
Qy 336 GGGCACTGCTTATCTTCAAGGTGACGAGCACTTCCCGCGGACGAGCC 395
Db 406 GGGTTCATATCAAGGCTGACATGAAGCTGATGGAAGCGGTTTCCCTGACGCGCC 465
Qy 396 CGTATGAAGAACAGAGCGCGCTGGAGGCCAGC 432
Db 466 AGTCATGACCAAGCATGTTGTCACAGAGCGGCTGC 502
|||||

RESULT 6
LOCUS CC649516 944 bp DNA linear GSS 19-JUN-2003
DEFINITION OGLC16TH ZM 0.7.1.5_KB Zea mays genomic clone ZM6BWA0322D08,
genomic survey sequence.
ACCESSION CC649516
VERSION CC649516.1 GI:32052184
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
1 (bases 1 to 944)
White, J.A., Quackenbush, J., Van Aken, S., Utecht, T.,
Citek, R.W., Nunberg, C.M., Budiman, M.A., Bedell, J.A., Kohlring, T.,
Resnick, A., Fraser, C.M., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGLC16TV
Contact: Cathy Whitelaw
TIGR

REFERENCE 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..944
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM6BWA0322D08"
/clone_lib="ZM 0.7.1.5_KB"
/note="Vector: pBCKS-; Site 1: HindIII; 0.7-1.5 kb
methylation filtered genomic DNA library"

FEATURES

source
1..944
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM6BWA0322D08"
/clone_lib="ZM 0.7.1.5_KB"
/note="Vector: pBCKS-; Site 1: HindIII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 9.7%; Score 66.8; DB 29; Length 944;
Best Local Similarity 45.7%; Pred. No. 4.2;
Matches 273; Conservative 0; Mismatches 322; Indels 3; Gaps 1;
30 GGCATCAAGATGTACATGAGGACCGGTGAACGCACTTCAATGTCGAGAGGCGCA 89
|||||

Db 129 GGGCTCGCCCTGTTCTAGCGCGCGCCCGCGCTGCTGTCGACCCCTCGGCGCGCAC 188
Qy 90 GGGCGACGGCAACCCCTTCGCGCGGCAACCCAGACATGAGATCCAGCTGACCGAGGCGC 149
Db 189 CGGCGTGGCGCAACGGCGGACTGCTCTACTTCAACGGCGCGCTCTGCGCATGTCGAGAA 248
Qy 150 CCCCCTGCTTCCTGCTTGCATCTGTGCCCCCTGCTGCTGAGTACGGACGACGACCTT 209
Db 249 CAGACTGCGGTACAGTCCGCGCTCGCGGACGACGCGGACCTGAGACCGTCCGCGCTA 308
Qy 210 CCGGACCAACCGCGGAGATCCCGCATCTTCTTCAAGCAGAGCTTCCCGAGGCTTCAC 269
Db 309 CCACTTCAGCGGCACTCGCTGCCCATGATCGGCAACCCCAACTGAGACCCGCGCAC 368
Qy 270 CTGGAGAGAGAACCAACCACTTACAGAGACGGCGGATCTTACCGCCACAGGACACAG 329
Db 369 CGGGAGCTGCAACGCGCTCAGCTACGAGGTCTGACAGAGCCCTTACTTCAAGTCTTCA 428
Qy 330 CCTGAGAGGCACTGCTGA--TCTAACAGGTGAAGTGCACGAGCACTTCCCGC 386
Db 429 CTTCAGGCGCGACGCGACCAAGTCCGACGACGTGAGATCCGCTGCGCCAGCCACAT 488
Qy 387 CCAAGGCGCCCTGATGAAGAACAGAGCGCGCTGGAGGCCACGACGAGGTGTGTA 446
Db 489 GATCCAGACTTGCATCAACGAGAACTTGTGCTGTGCTCCGACCAACAGGTGTGTT 548
Qy 447 CCGCGAAGCGGCGTGTGTCGCGCGGAACTGATGCGCTGAGGTGGAGGACCGGCA 506
Db 549 CAAAGCTGCAAGAGATCTGCGCGCGGCTGCGCGCTGCTGACAGAGGACGACGTC 608
Qy 507 CCTGATCTGCAACCACTACACAGCTACCGGACCAAGAGCGGCTGCGCCCTGACAT 566
Db 609 GCGCTTCGCGGTCTCCCGAAGCGCGCGGACGCGTGGAGATGCGGTGTGACGT 668
Qy 567 GCGCGCTTCATCTTCAACGACATCCGCGCTTCCAGTGTGCGGAGAAAGACGAG 624
Db 669 GCGGACTGCTTCTGCTTCCACTGTGGAACGCGTGGAGAGACGAGCGACGCGCGAG 726

RESULT 7
LOCUS CP864358 700 bp mRNA linear EST 31-OCT-2003
DEFINITION pzs009xp17f USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation_szs Phytophthora sojae cDNA clone szs009p17 5, mRNA sequence.
ACCESSION CP864358
VERSION CP864358.1 GI:38118984
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
REFERENCE 1 (bases 1 to 700)
AUTHORS Tyler, B.
TITLE Unpublished (2003)
JOURNAL Contact: Tyler B
COMMENT Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmyler@vt.edu
PCR PRIMERS
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 009 row: P column: 17
Seq primer: BK reverse primer
High quality sequence stop: 700.
Location/Qualifiers
1..700
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"

FEATURES
source

/clone="szs009p17"
/issue_type="Zoospores"
/cell_line="P6497"
/dev_stage="Free swimming"
/clone_1ib="USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation_szs"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Query Match 9.7%; Score 66.4; DB 14; Length 700;
Best Local Similarity 45.7%; Pred. No. 4.5; Mismatches 316; Indels 3; Gaps 1;
Matches 269; Conservative 0;

Qy 66 CCACTACTTCAAGTGCAGAGGCGAGGCGACCGCAACCCCTTTCGCGGACCCAGACAT 125
Db 103 CGGCAACCAACAGAGCAAGGCTGATGTCATGACCCGCGCGGACGACATCCACCGT 162
Qy 126 GAGATTCAGTGAACCGAGGCGCCCTGCTTGCCTTGCATATCTGGCCCTTG 185
Db 163 GAG---CTCCGTGCGCACGCTGCGCGCGCACGACGACAGTCCGCCACAGCCCGCG 219
Qy 186 CTGCAAGTACGCGACAGGACCTTGTGCAACACACCGCGAGATCCCGACTTCTCA 245
Db 220 CCTTCCACGGATCAAGTGCAGAGGTCTTCCCAAGAGCGCATGACGTATCTA 279
Qy 246 GCAAGACTTCCCGAGGCGCTTCACTGAGAGAAACCAACCTTACGAGACGCGCAT 305
Db 280 CCACTACTTCAAGAGCGGATCTGAGTGCAGGCGCGCTGATGAGGTGTGCGGCTCGG 339
Qy 306 CCTGACCGCCCAACAGACACGACCTGAGAGGCAACTGCTGATCTACAGGTGAAGT 365
Db 340 CCTCGAGGACTCAGCAACTGCTGTCAACACGAGACCATCAAGTACGAGACATCCC 399
Qy 366 GCAAGGCAACCACTTCCCGCGCGAGCGCGCTGATGAAGAACAGAGGCGCGCTGGA 425
Db 400 GAGTTCCTCCGCTTCCAGGAGCGACGCGGAGCTGTCTTGGGACCTGGA 459
Qy 426 GCCAGACCAAGAGTGTATCCCGAGAACGCGGTGTGTCGCGCGGACGATGAGC 485
Db 460 GGGCTTCGCGCTGTGTGATGCGGCGCGCTTCCACTGCTACAGAGGCTATGCGCATG 519
Qy 486 CCTGAAGTGGCGACCGGACCTGATCTGCCACCACTACACGACTACCGAGCAAGAA 545
Db 520 CAAAGCGGCGTGCCTATCGGTCAATCTCTCCGCGCATCAAGTACTTCTGCTGAC 579
Qy 546 GCGCGTGGCGGCGCTGACATCCCGGCTTCCACTTCAACCGACATCCGCTCAAGTGT 605
Db 580 CAACGCGCGGCGGCGCTGAACCCGACCTTCAACGTCGAGCGACGTATGATGAACGA 639
Qy 606 GCGAAGAAAGAGACGAGTACTTGCAGCTGTACGAGCGCACCGTGGC 653
Db 640 CCACTGAACGTGCGCGGCTGTGAGTACGACCCCGCTCATCGGCGC 687

RESULT 8
LOCUS BZ422085 624 bp DNA linear GSS 10-DEC-2002
DEFINITION id49d07.b1 WGS-Shicolorf (DH5a methyl filtered) Sorghum bicolor genomic clone id49d07 5', genomic survey sequence.
ACCESSION BZ422085
VERSION BZ422085.1 GI:26370029
KEYWORDS GSS.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE 1 (bases 1 to 624)
AUTHORS Rabinowicz, P.D., O'Shaughnessy, A.L., Bailja, V., Dedhia, N., Katzenburger, F., King, L., Muller, B., Muller, S., Nascimben, L., Zutter, T., Palmer, L., McComb, W.R., and Martienssen, R.A.
TITLE Genomic shotgun sequences from Sorghum bicolor (methyl-filtered) Unpublished (2002)

Fri Jul 16 08:44:56 2004

us-09-976-673-11.tbl

Page 6

COMMENT

Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100 Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Place: 1d49 row: d column: 07
Seq primer: -21M13UnivFwd
Class: Shotgun

FEATURES

High quality sequence stop: 624.
Location/Qualifiers

source

1..624
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/db_xref="taxon:4558"
/clone="1d49d07"
/lab_host="DH5a"
/note="Site 1: Xba I; Site 2: Xba I; The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using Sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, b/g reads in pUC19). The same ligation was transformed into DH5a."

ORIGIN

Query Match 9.6%; Score 65.8; DB 28; Length 624;
Best Local Similarity 48.1%; Pred. No. 5.4;
Matches 187; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 79 TCGAGGGGAGGGGAGCGGCAACCTTGGCCGGGACCCAGAGATGAGATCCAGT 138
DB 58 TTCACGTCAAGTGTGTGGGAGCGACAGATGCGATCCACCGGTAGATCCGGGG 117
QY 139 ACCGAGGCGCCCCCTGCGCTTGTGACATCTTGCCCTCTGTGGAGTACGGC 198
DB 118 CCGCGAGAGCGCCCAACAGGAGCGGTGCAATCAGAGGCTCTTCAAGTGTGGGTC 177
QY 199 AGCAGAGCTTGTGCAACAACCGCGAGATCCCGACTTCTTCAAGCAGAGTTTCCC 258
DB 178 AGGAGTGGCGGTGGCCACCGGAGAGATGCTGCTCCGTGGCCCGGCGCTTCGAC 237
QY 259 GAGGCTTCACTGGGAGAGAACCAACCTACGAGAGAGCGGCGATCTGACCGCCAC 318
DB 238 GTGTGTGTGTCCGGGGTGTGTGGGCGCCCGGACAGGATCAGGTGGAGCTTCGGC 297
QY 319 CAGGACACGAGCTGAGAGGCACTGCTGATCTTCAAGTGAAGTGCACCGGACAAAC 378
DB 298 AGGTACCGCGGAGAGAGAGTCCCGGCTGCGGTGCGCACTGACCTTGGCGGC 357
QY 379 TTCCCCCGGAGCGCCCGTGTATGAAGAACAGAGCGGCTGGAGAGCCAGACCGAG 438
DB 358 ACCTCAACGCGGTGCGATCAAGACGTGGCGGCGGACCTGACCAACCGCGCTGACC 417
QY 439 GTGTGTATCCCCGAAACGCGGTGTGTG 467
DB 418 GGCCTGCTCTTGGAGGACATGTCATGAG 446

RESULT 9

CC424389

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

982 bp DNA linear GSS 19-MAY-2003

PHUNK57TB ZM 0.6-1.0_KB Zea mays genomic clone ZM8BTA491109,

genomic survey sequence.

CC424389.1 GI:30904479

GSS.

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 982)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2003)
Other GSSs: PUHNS31ND
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES

source

1..982
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM8BTA491109"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

ORIGIN

Query Match 9.6%; Score 65.8; DB 28; Length 982;
Best Local Similarity 46.2%; Pred. No. 5.9;
Matches 224; Conservative 0; Mismatches 337; Indels 6; Gaps 2;

QY 28 ATGCCCATCAAGATCTCATAGAGGACCGGTAAGCGCCACTTCAATGCGAAGGC 87
DB 263 AAGGGGTCTCATGTGCTGCGCCGACGTGTGATCAACACCGGTGCGGACTTAAGAGAC 322
QY 88 GAGGGGAGCGGCAACCCCTTTCGCGGCAACCAAGAGATGAGATTCACGTGACCGAGGC 147
DB 323 GGCCTCGGCACTTCACTGCTGTTCGAGGGGAGGACCGCCGACAGCGCTGACTGGGGC 382
QY 148 GCGCCCTGCGCTTGTGCTTGCATCTCTG---GCCCTCTGTGAGTACGAGCAGG 204
DB 383 CCGGACATGATCTGACAGAGACGACGACGATCTCAACGGGCGGGGACCGGACACG 442
QY 205 ACTTGTGACACGACACGCGGAGATCCCGACTTCTTCAAGCAGAGCTTCCCGAGGGC 264
DB 443 GGGGCGGACTTGTGCGCGCGCGCCGACATGACACCTCAACCGCGCTGTGACAGAGAG 502
QY 265 TTCACTTGGAGAGAAACCAACCTTCAAGAGACGCGGCGATCTGACCGCCACAGAGAC 324
DB 503 CTCTGGACTGGCTCAATGAGCTCAAGTCCGACTTGTGACCGGTGAGCGCTTGAC 562
QY 325 ACCAGCTGAGAGGCAACTGCTGATCTCAAGGTGAAGGTGCAACGCGACCAACTTCCC 384
DB 563 TTTCGCAAGGGCTTACTCGCGCGCGGCTGCGCAAGGTGTCAGTCAAGACGCGCCACCC 622
QY 385 GCGGAGCGCCCGTGTATGAAGAACAAAGAGCGGCGCTGGAGCCCAACGAGTGTG 444
DB 623 TTGTGTGTGCGCGAGATATGAGAG---TCCCTTCACTTCAAGCGGCAACGCGAGCGCTTC 679
QY 445 TACCCCGAGAGACGCGCGTGTGTGCGCGCGGAACTGATAGCGCTTGAAGGTGGCGACCG 504
DB 680 AGCAACGAGAGCGCGGACAGAGAGTGTGTCAACTGGCGCGAGCGGTGTGGCGCGCC 739
QY 505 CACTGATCTGCAACCACTACCAAGCTTCCGAGCAAGAAAGCGCGCTGCGCGCTTGAAC 564
DB 740 GCGCGGCGTTCGACTTCAACCAAGGCGGTGTGACAGGCGCGCGCTTCAAGGCGAGCTG 799
QY 565 ATGCCCGGCTTCACTTCAACGACATCCGCGCTTCAAGTGTGCGGAGAAAGAGAGAGAG 624
DB 800 TGGCCCATGAAGAGCGGCAACGCGCAAGGCGCGCGGAGTGTGCTGCTGCGCGAGAGAG 859

QY 625 TACTTCGAGCTGACGAGCCGACGCTGGCCCCGGTACA 661
DB 860 GCCGTACCGTTCGTGACAAACGACGACCGGCTCCA 896

RESULT 10
LOCUS CB651092/c 794 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNB1018.r OSJNB Oryza sativa (japonica cultivar-group) cDNA
clone OSJNB15018 3', mRNA sequence.
ACCESSION CB651092
VERSION CB651092.1 GI:29646085
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriocaridaceae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 794)
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,B.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 440A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3867
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gga aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 15 row: 0 column: 18
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1. '794
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB15018"
/issue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

ORIGIN
Query Match 9.5%; Score 65.2; DB 14; Length 794;
Best Local Similarity 49.7%; Pred. No. 6.9;
Matches 197; Conservative 0; Mismatches 193; Indels 6; Gaps 1;

QY 74 TCAAGTGCAGGGGCGGCGGAGGCAACCCCTTCGCGCCGACCGAGCATGAGAATCC 133
DB 793 TATATGCTGTGTCGCGGACGTCGCGACACGCGCGCTGCGCAAGACGACAGTACC 734

QY 134 ACGTGCACGAGGGGCGCCCTTCGCTTCGACATCTCGGCCCTTCGTGCGAGT 193
DB 733 GCAAGCCGAGCTTCTACACCATGAGGCGCGGCTACGTAACCGTCCCGCGCTGACC 674

QY 194 ACGGACGACGAGCTTCTGTCACACACGCGCGAGATCCCGACTTCTTCAACAGAGT 253
DB 673 CGGGCTCGTCTTACGACTTCACACCAAGACTACGCTTCTGCGGCTCGGCA 614

QY 254 TCCCGAGGGCTTCACTCGGAGAGAACCAACCACTACGAGACGCGCGCATCTTACCG 313
DB 613 TCGGCGACGAGCGCTGAAGAGATTAACCAACGCGCGCTTCTCTGCGCAAGCTCAAG 554

QY 314 CCCACGAGACACCAAGCTTGAGGGCACTGCTTACTTACAAGTGAAGTGCACGCA 373
DB 553 CCATCACCGAGGCG-----GGAGCTCACTACCCGCTGGTGGTGAAGCTGTCC 500

QY 374 CCAACTTCCCCGCGACGCGCCCTTGATGAAGAACAGCGCGCTGGAGCCAGCA 433
DB 499 AGCCGATCACCGTTCACCGCATCTGTACCAACGTCGGGAAGGCCAATCCGTACCG 440

QY 434 CCGAGGTGTGTACCCCGGAACGCGCTGTGCG 469
DB 433 CCGTGTGACATGCCCAAGAACGTGCGGTACCG 404

RESULT 11
LOCUS BU626888/c 697 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-FTO-bhn-d-01-0-UI.81 NCI CGAP_FTO Homo sapiens cDNA clone
UI-H-FTO-bhn-d-01-0-UI 3', mRNA sequence.
ACCESSION BU626888
VERSION BU626888.1 GI:23293103
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Robeff-Pamela, U of I
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=yes
Location/Qualifiers
1. '697
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FTO-bhn-d-01-0-UI"
/issue_type="Alveolar Macrophage"
/dev_stage="Adult"
/lab_host="NCI CGAP FTO"
/clone_lib="NCI CGAP FTO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP_FTO is a cDNA library constructed from a pool of
81 RNA samples from Alveolar Macrophages challenged with
different treatments. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GGCCATGCCG. The cell line
was provided by Gary W. Humminglake from the University of
Iowa.
TAG TISSUE=Human Lung Alveolar Macrophage
TAG LIB=UI-H-FTO
TAG_SEQ=GGCCATGCCG"

ORIGIN
Query Match 9.4%; Score 64.8; DB 13; Length 697;
Best Local Similarity 52.6%; Pred. No. 7.6;

Matches 141; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 214 CACCACACCGCCGAGATCCCGCTTCTTCAAGACAGAGCTTCCCGAGGCTTCACTTGG 273
 Db 671 CCGACCAATGAAAGACGACGCTTCTTCAAGTCCCGCATGCGCGAAGGTAACTGTCAG 612

QY 274 GAGAGAACCAACCACTACGAGGACGCGGCACTCCGACCGCCCAACGACCAAGCTTG 333
 Db 611 GAGCGACCACTTCTTCAAGACGACGCACTCAAGACCGCGCGGAGGTGAATTC 552

QY 334 GAGGCAACCTGCTTCTTCAAGGAGGACGCGGCACTTCCCGCGGCAAGCTTG 393
 Db 551 GAGGCGACACCTGCTTCAAGGAGGACGCGGCACTTCAAGGAGGACGCGG 492

QY 394 CCGGTGATGAAAGAACAGAGCGCGGCTGGAGCCGACCGAGGTGCTTACCCGAG 453
 Db 491 AACATCTGGGGGCAAGCTGAGTCAACTCAACAGCAGCAAGCTATATATGAGCC 432

QY 454 AACGCGCTGTGCGCGCGGCAACTGA 481
 Db 431 GACAAGCAAGAAAGCGCATCAAGTGA 404

RESULT 12
 CK199961 821 bp mRNA linear EST 08-DEC-2003
 LOCUS FGAS008468 Triticum aestivum FGAS: Library 3 Gate 6 Triticum
 DEFINITION aestivum cDNA, mRNA sequence.
 ACCESSION CK199961
 VERSION CK199961.1 GI:39562351
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Pooideae; Triticeae; Triticum.

REFERENCE
 1 Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Htycan, L.D., Laroche, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
 Penikett, C., Roach, J.T., and Searles, F.

TITLE
 JOURNAL
 COMMENT
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops
 Unpublished (2003)
 Contact: Wm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas.estecs.usask.ca

FEATURES
 source
 1..821
 Location/Qualifiers
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_lib="Triticum aestivum FGAS: Library 3 Gate 6"
 /note="Organ: Root; Vector: pCMV.SPORT6; Root tissue from
 control, cold-acclimated and salt stressed wheat cultivar
 Norstar. 7 mRNA populations were combined before
 constructing the library; 7 day non-acclimated roots, 1,
 23, and 53 days cold-acclimated at 4C, and 30 minutes, 3
 hours and 6 hours treated roots with 200mM NaCl.
 Non-acclimated and cold-acclimated plants were grown in
 vermiculite while salt stressed plants were grown in
 hydroponically. First strand synthesis in this library was
 done in the presence of methylated dCTP thereby protecting

ORIGIN from internal cleavage with NctI."

Query Match 9.4%; Score 64.6; DB 14; Length 821;
 Best Local Similarity 48.1%; Fred. No. 8.4; 229; Indels 3; Gaps 1;
 Matches 215; Conservative 0; Mismatches 229; Indels 3; Gaps 1;

QY 104 CTTTGGCGGACACCCAGAGATGAGATTCACGATGACCGAGGCGCCCTTCCCTTGG 163
 Db 293 CCCCCCGGCAACACACCGGCCCCCGGCGCGGACCGGACGCTTCTTCCCATCCCGG 352

QY 164 CTTTGACATCTGAGCCCTTCTGCTGAGTACCGGACGAGACCTTGTGTACCAACCG 223
 Db 353 CGCCGCAACAAACAGAGCGCGGAGTTCGCGCGCCGACAGATGACGAGGCTTCAAGGGG 412

QY 224 CCGAGATCCCGACTTCTTCAAGCAGAGCTTCCCGAGGCTTCACTGGAGAGAACCA 283
 Db 413 CCGGCTCATGAGTTCGCGGAGAGCTTGGGCGCACACCAACCAAG---GCCG 469

QY 284 CCACTTACAGAGAGCGGCTTCTGACCGCCCAACAGACAGACCTGAGAGGCAACT 343
 Db 470 ACCGCTACAAACAGTACGCTTCCGACGCGCAAGCCCTTACGCAAGCCCAAG 529

QY 344 GCGTATCTACAGAGTGAAGTGCACGAGCACTTCCCGGAGCGCCCTGATGA 403
 Db 530 CCAACGCGATGATGAGTGCACGAGCCCAACCAACCAACATGAAACAGCTTGGGGCTCA 589

QY 404 AGAACAAGAGCGGCGCTGGAGCCGACGACCGAGTGTATCCCGAGAACGCGCTGC 463
 Db 590 ACAAGATGCTGCGCGGCGGCGGAGCTTCAAGCAAGCGGCAACGAGGGGTGA 649

QY 464 TGTGGCGCGGAGAGTATGAGCTTGAAGTGTGAGCGACCGGACCTGATCCACCACT 523
 Db 650 AAGGCTATCTTCAACAAATTCATGCGAGGCGCGGCAACAAAGCAGCAGCCACT 709

QY 524 ACACAGCTTACCGGACGCAAGAGCGG 550
 Db 710 TCAGCGGCTTACGCGGCGGCAAGAGGCG 736

RESULT 13
 B1366976 613 bp mRNA linear EST 26-AUG-2003
 LOCUS BFL26 001679 Amphioxus 26hr cDNA library (Name convention: BFL26 or
 MPMG531) Branchiostoma floridae cDNA clone MPMG531IN0799 5', mRNA
 DEFINITION sequence.

ACCESSION B1366976
 VERSION B1366976
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 Branchiostoma.

REFERENCE
 1 (bases 1 to 613)
 Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
 Herwig, R., Vingron, M., and Lehrach, H.

TITLE
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes

JOURNAL
 MEDLINE
 PUBMED
 Genome Res. 13 (6A), 1056-1066 (2003)
 22683279
 12799346

COMMENT
 Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Inmestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1128
 Fax: +49 30 8413 1128

Email: panopoulou@gen.mpg.de
 The library was characterized by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridization pattern with a
 battery of 200 oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
 FORWARD: 5' CCCGAGCTTACCTTATGCTCCGGCTCG 3' (M13RSP)
 BACKWARD: 5' GCTATTACGCCAGCTGGCGAAGGGGAGTGTG 3' (M13FSP)
 Insert Length: 1 Std Error: 0.00
 Seq primer: 5'-CCGCTCCGGAATTCGCGGT-3' pSport3/86
 High quality sequence stop: 613.

FEATURES

source

```

1..613
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMG531N0799"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMG531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-GACTACTTCTAGATCGGAGCGGCGCC (7)15-3' and a SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."
```

ORIGIN

Query Match

Best Local Similarity 47.6%; Pred. No. 9.1; Length 613;
 Matches 189; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```

QY 36 CAAGATGTACATGAGAGGACCGGTGAACGGCACTTCAAGTGCAGAGGCGAGGGCGCA 95
DB 88 CGACATCCACCTTCAAGGCTCCATCAACGGCCACGAGTTGACATGTGGGGGAGAGAA 147
QY 96 CCGCAACCCCTTGGCGGACCCAGAGCATGAAATCCAGTACCGAGGGCGCCCTTGA 155
DB 148 AGCGACCGGAAAGCGCGCTCGGTGATGACAGGGAATCCACCAAGGGTGCCTTGA 207
QY 156 GCCCTTGCCTTGCATCTCTGAGCCCTGCTGCTGAGTACGAGCAGAGCACTTGTGCA 215
DB 208 GTTCTCTCCCTTATGATATATCCCACTCGGCTACGGGTACTACAGTACTCCCTTA 267
QY 216 CCACACCGCGGAGATCCCGACTTCTTCAAGCAGAGCTTCCCGAGGGCTTCACTGGGA 275
DB 268 CCGGAGCGGACCTCGCTTTTCAAGGCTTCATGTTGAAGATCGGGGTATGCACTTA 327
QY 276 GAGAACCACTTACAGAGAGCGGCGCATCTGACCGCCACCAAGACACCAAGCTTGA 335
DB 328 CCGCGTGTTCGATTTGAAGACGAGGCAAGCTGACTGACGAGTTTAAGTACTCTTACA 387
QY 336 GGGCAACCTGCTGATCTTCAAGGTGAAGTGCAGGACCACTTCCCGCGGAGGGCC 395
DB 388 GGGTTCCTCATATTAAGCGGACATGAAGCTGATGGAGAGCGGTTTCCCTGATGACGCC 447
QY 396 CGTGATGAAGAACAGAGCGCGGCTGGAGGCCAGC 432
DB 448 AGTCATGACACGACGAGTTGTGACCAAGAGCGGCTGC 484
```

RESULT 14

BI386699

LOCUS BI386699 643 bp mRNA linear EST 26-AUG-2003
 DEFINITION BFL26 001402 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMG531) Branchiostoma floridae cDNA clone MPMG531I1370 5', mRNA

ACCESSION

BI386699
 BI386699
 VERSION BI386699.1 GI:30921604

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

EST.
 Branchiostoma floridae (Florida lancelet)
 Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
 1 (bases 1 to 643)
 Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J., Herzig,R., Vitzthum,M. and Lehrach,H.
 New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
 Genomes Res. 13 (6A), 1056-1066 (2003)
 22683279
 12799346
 Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr.63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoulou@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
 FORWARD: 5' CCCGAGCTTACCTTATGCTCCGGCTCG 3' (M13RSP)
 BACKWARD: 5' GCTATTACGCCAGCTGGCGAAGGGGAGTGTG 3' (M13FSP)
 Insert Length: 1 Std Error: 0.00
 Seq primer: 5'-CCGCTCCGGAATTCGCGGT-3' pSport3/86
 High quality sequence stop: 643.

FEATURES

source

```

1..643
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMG531I1370"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMG531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-GACTACTTCTAGATCGGAGCGGCGCC (7)15-3' and a SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."
```

ORIGIN

Query Match

Best Local Similarity 47.6%; Pred. No. 9.2; Length 643;
 Matches 189; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

```

QY 36 CAAGATGTACATGAGAGGACCGGTGAACGGCACTTCAAGTGCAGAGGCGAGGGCGCA 95
DB 77 CGACATCCACCTTCAAGGCTCCATCAACGGCCACGAGTTGACATGTGGGGGAGAGAA 136
QY 96 CCGCAACCCCTTGGCGGACCCAGAGCATGAAATCCAGTACCGAGGCGCGCCCTT 155
DB 137 AGCGACCGGAAAGCGCGCTCGGTGATGACAGGGAATCCACCAAGGATCCCTTGA 196
QY 156 GCCCTTGCCTTGCATCTCTGAGCCCTGCTGCTGAGTACGAGCAGAGCACTTGTGCA 215
DB 197 GTTCTCTCCCTTATGATATATCCCACTCGGGTATGAGTACTACAGTACTCCCTTA 256
```

Qy	216	CCACACCGCCGAGATCCCCGACTTCTCTAAGAGAGAGTTCCCGAGGGCTTCACTGGGA	275
Db	257	CCCGGACGGACCCCTCGCTTTTCCAGGCTCCATGTTTGAAGAGATCGGGGTATGACGTCTA	316
Qy	276	GAGAACCAACCACTTACAGGAGCGCGCGCATCTCTGACCGCCACACAGGACACCAAGCTGGA	335
Db	317	CCGCGCTGTTCGACTTTGAGAGCGGAGGCAAGCTGACTACCGAGTTTAAATACTCCTAAGA	376
Qy	336	GGGCAACCTGCTGATCTTACAAGGTGAAGGTGCAAGGCACCAATTTCCCGCGGACGGGCC	395
Db	377	GGGTTCCTCATATCAAGGCCGACATGAAAGTGAAGGAAAGGGTTTCCCTGACGAGGGCC	436
Qy	396	CGGTATGAAGAACAAAGACGGCGGCTGGGAGGCCAGC	432
Db	437	AGTCATGACCAAGCAATTTGTGCAACCAAGGACGGCTGC	473

RESULT 15	
CNS02EOD/c	
LOCUS	681 bp linear GSS 01-SEP-2000
DEFINITION	CNS02EOD Tetradon nigroviridis genome survey sequence T7 end of clone 252H4 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION	AF193990
VERSION	AF193990.1
KEYWORDS	GI:7832096
SOURCE	GSS; genome survey sequence.
ORGANISM	Tetradon nigroviridis Tetradon nigroviridis

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBLISHED	REFERENCE	AUTHORS
1	Reest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Berron, A., Fikames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W., and Weissenbach, J.	Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence	Nat. Genet. 25 (2), 235-238 (2000)	20296633	10835645	2	Reest Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C.,

TITLE Characterization and repeat analysis of the compact genome of the freshwater puffball *Tetradodon nigroviridis*

JOURNAL MEDLINE 20359837

PMID 10892143

REFERENCE 3 (bases 1 to 681)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-MAR-2000) Genoscope - Centre National de Séquençage / Institut National de la Santé et de la Médecine / Centre National de Séquençage

COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon .
FEATURES	location/Qualifiers
SOURCE	1. 681 Tetradon nigroviridis

ORIGIN

Query Match 9.3%; Score 64; DB 29; Length 681;
Best Local Similarity 41.9%; Pred. No. 9.9;
Matches 176; Conservative 27; Mismatches 217; Indels 0;
Gaps 0.

OY	18	GAAAGAAAGTATGCGATCAATGATATGATGAAAGGCAACCGTGAACGGCCATCTACTCA	77
Db	433	GGAGACRATTTTATTATTAATTTTGTCDACRCAACACACAAACAAAACRACACCTA	376
OY	78	GTGCGAAGGCGAAGGCGGACCGCCCTTGCCGGCACCACAGCATGAGTAATCCAGT	13
Db	373	CAACAAATAACAAACACCGCGCGGACAAACACGAGACRACGACACACAAACGACCA	31
OY	138	GACCGAAGGCGCCCCCTGCGCTTGCGCTTGCGATCTTGCGCCCTGCTGCGAGTACGG	19
Db	313	CGGCGACGACGACGACCAACRACGGCGAGCGACGACGACGACGACACAAACACGG	25
OY	198	CAGCAGACCTTGCTGTGACCAACCGCGAGATCCCGCATTTCTTCAAGCAGAGTCC	25
Db	253	CACACACGACGGCGAC	19
OY	258	CGAAGGCTTCACTTGGAAGAAACACCACTTACGAGACGGCGGATCTCTACCGGCCA	31
Db	193	CGGACACAAACAAACACRACRACAAAACRACACACRACGACACACACRACACAC	13
OY	318	CCAGACACACGACCTGAGGGCAACCTGCTGATCTTCAAGGTGAAGGTGACCGCCACAA	37
Db	133	CAACAAAAACAAACAAACAAACAAACAAACAAACRACRACRACRACRACRACRAC	74
OY	378	CTTCCCCCGCAGCGCCCGGTGATGATGAACAAGCGCGCGGCTGCGGAGCCACGACGA	43
Db	73	AAACACACRACACCGCAACACASACACACACACACACACACACACACACACACAC	14

Search completed: July 15, 2004, 14:05:20
Job time : 2894 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2004, 10:05:08 ; Search time 395 Seconds
(without alignments)
7388.638 Million cell updates/sec

Title: US-09-976-673-11

Perfect score: 687
Sequence: 1 atcgctgagcgccctgctgaa.....tgcccgagaagggccaactga 687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980a:*
2: geneseqn1990a:*
3: geneseqn2000a:*
4: geneseqn2001a:*
5: geneseqn2001b:*
6: geneseqn2002a:*
7: geneseqn2003a:*
8: geneseqn2003b:*
9: geneseqn2003c:*
10: geneseqn2004a:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	100.0	687	6	ABL41172 H. crista
2	671	97.7	687	6	ABL41173 H. crista
3	671	97.7	1386	6	ABL41174 H. crista
4	671	97.7	1396	6	ABL41175 H. crista
5	669.4	97.4	1424	6	ABL41175 Cr-449-ta
6	669.4	97.4	1424	7	ACa61025 DNA encod
7	446.6	65.0	1376	7	ACa61026 DNA encod
8	441.2	64.2	681	6	ABL41170 H. crista
9	439.6	64.0	681	6	ABL41171 H. crista
10	438.2	63.8	684	6	ABL41180 H. crista
11	436.6	63.6	684	6	ABL41169 H. crista
12	436.6	63.6	760	7	ABA00802 Wild type
13	436.6	63.6	760	7	ABA00806 Multiple
14	435	63.3	910	6	ABL41167 H. crista
15	435	63.3	910	6	ABL41182 H. crista
16	431.8	62.9	760	7	ABA00805 Multiple
17	431.8	62.9	760	7	ABA00804 Chromopro
18	431.8	62.9	760	7	ABA00803 Chromopro
19	431.8	62.9	908	6	ABL41168 H. crista
20	429.2	62.5	680	6	ABL41181 H. crista
21	422.6	61.5	686	3	AAA50885 A. sulcat
22	417.8	60.8	707	6	AAD46286 Anemonia
23	417.8	60.8	1398	7	ACA61028 DNA encod

24	415.6	60.5	1404	7	ACA61027 DNA encod
25	415.2	60.4	699	6	AAD46287 Anemonia
26	362.6	52.8	681	6	ABA00245 C. gigant
27	361.2	52.6	835	6	ABA00244 C. gigant
28	306	44.5	678	9	ADC24132 Discosoma
29	298.8	43.5	681	4	AAH47656 Anthozoan
30	298.2	43.4	681	9	ADC24130 Discosoma
31	297.4	43.3	723	7	ABD22476 Mammalian
32	292.4	42.6	678	6	AAD46278 Discosoma
33	292.4	42.6	678	6	AAD28208 Discosoma
34	292.4	42.6	678	6	AAD28207 Discosoma
35	292.4	42.6	695	3	AAA48743 Humanised
36	291.8	42.5	678	9	ADC24128 Discosoma
37	291.4	42.4	681	9	ADC24134 Discosoma
38	291.4	42.4	4692	6	AAL47954 Modified
39	291.4	42.4	4692	7	ACC44640 Vector PD
40	291.4	42.4	7910	4	AAD09979 PBIT(DHSP
41	291.4	42.4	9320	6	ABSS6664 Plasmid P
42	290.8	42.3	6893	9	AAD24111 Proviral
43	289.8	42.2	5436	4	AAD10003 Plasmid P
44	289.2	42.1	678	6	AAD46282 Discosoma
45	289.2	42.1	678	6	AAD28209 Discosoma

ALIGNMENTS

RESULT 1	ABL41172	standard, cDNA, 687 BP.
AC	ABL41172;	
XX		
DT	12-AUG-2002 (first entry)	
XX		
DE	H. crista fluorescent protein mutant FP10-ctrl encoding cDNA.	
XX		
KW	Stichodactylidae; chromoprotein; fluorescent; anthozoan; food; FRET;	
KM	colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;	
KW	fluorescence resonance energy transfer; gene expression; mutant; ss.	
XX		
OS	Heteractis crista.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..687
FT		/*tag= a
FT		/product= "mutant fluorescent protein"
XX		
PN	WO200230965-A2.	
XX		
PD	18-APR-2002.	
XX		
PF	12-OCT-2001; 2001WO-US032080.	
XX		
PR	12-OCT-2000; 2000US-0240018P.	
PR	16-JUL-2001; 2001US-0306131P.	
XX		
PA	(CLON-) CLONTECH LAB INC.	
XX		
PI	Lukyanov SA, Fradkov AF, Lukyanov KA, Gurskaya NG;	
XX		
DR	WPI; 2002-444170/47.	
XX		
DR	P-PSDB; ABB07994.	
XX		
PT	Novel nucleic acid encoding Stichodactylidae chromoprotein and its	
PT	assays, markers in recombinant DNA applications and filters in	
PT	suncreens.	
XX		
PS	Claim 5; Fig 10; 81pp; English.	
XX		
XX		
CC	The invention relates to a nucleic acid present in other than its natural	
CC	environment and encoding an Stichodactylidae chromoprotein or its	

CC fluorescent mutant, where the fluorescent protein has an emission maximum
 CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
 CC useful in applications employing a chromo or fluorescent nucleic acid or
 CC protein. Recombinant vectors comprising the nucleic acid is useful for
 CC producing an Anthozoan chromo and/or fluorescent protein. The
 CC chromoproteins, and their fluorescent mutants are useful as colouring
 CC agents capable of imparting colour or pigment to a particular composition
 CC of matter. The chromoproteins can be incorporated into a variety of
 CC different compositions including food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g. animals and plants, and as labels in
 CC analyte detection assays, e.g. assays for biological analyses of interest
 CC (see AB41167 for a detailed description of the various uses of the
 CC chromoproteins). The present sequence represents the H. crista
 CC fluorescent protein mutant FP10-cr1 encoding cDNA
 XX

SQ Sequence 687 BP; 150 A; 232 C; 212 G; 93 T; 0 U; 0 Other;
 Query Match 100.0%; Score 687; DB 6; Length 687;
 Best Local Similarity 100.0%; Pred. No. 1e-111;
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGAGCGGCGCTGTGAAGAGAGATGCGCATCAAGATGATCATGAGAGGACCGT 60
 DB 1 ATGTGAGCGGCGCTGTGAAGAGAGATGCGCATCAAGATGATCATGAGAGGACCGT 60
 QY 61 AACGGCCACTACTTCAAGTGCAGAGGCGGAGACCGCAACCCCTTGCCTGGACCCG 120
 DB 61 AACGGCCACTACTTCAAGTGCAGAGGCGGAGACCGCAACCCCTTGCCTGGACCCG 120
 QY 121 AGCATGAGATTCACGATGACGAGGCGCCCTTGCCTTGCCTTGCATCTTCGCTC 180
 DB 121 AGCATGAGATTCACGATGACGAGGCGCCCTTGCCTTGCCTTGCATCTTCGCTC 180
 QY 181 CCTGCTCGAGTACGCGACAGACCTTCTGTGACACACCGCGCGAGATCCCGACTTC 240
 DB 181 CCTGCTCGAGTACGCGACAGACCTTCTGTGACACACCGCGCGAGATCCCGACTTC 240
 QY 241 TTCAAGAGAGACTTCCCGGAGGGCTTCACTGGGAGAGAACCAACACTACAGAGACGG 300
 DB 241 TTCAAGAGAGACTTCCCGGAGGGCTTCACTGGGAGAGAACCAACACTACAGAGACGG 300
 QY 301 GGCATCTGATCCGCCCAACAGACCAAGCTTGAAGGGCACTGCTGATTTACAAGTTC 360
 DB 301 GGCATCTGATCCGCCCAACAGACCAAGCTTGAAGGGCACTGCTGATTTACAAGTTC 360
 QY 421 TGGAGAGCCAGACCGAGTGTATCCCGAGAAACGGGCTGTGCGGCGGAAAGTTC 480
 DB 421 TGGAGAGCCAGACCGAGTGTATCCCGAGAAACGGGCTGTGCGGCGGAAAGTTC 480
 QY 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATGTGCAACACTACACAGTTACCGAGG 540
 DB 481 ATGGCCCTGAAGGTGGGCGACCGGACCTGATGTGCAACACTACACAGTTACCGAGG 540
 QY 541 AAGAAAGCCGTGCGCGCTGACCAATGCGGCTTCACTTCAACGCAATCCGCTCCAG 600
 DB 541 AAGAAAGCCGTGCGCGCTGACCAATGCGGCTTCACTTCAACGCAATCCGCTCCAG 600
 QY 601 ATGCTTGGGAAAGAAAGAGACGATCTTGAAGCTGTACGAGGCGAGCGTGGCCCGGTAC 660
 DB 601 ATGCTTGGGAAAGAAAGAGACGATCTTGAAGCTGTACGAGGCGAGCGTGGCCCGGTAC 660
 QY 661 AGCGACTGCGCGAGAGGCCCAACTGA 687
 DB 661 AGCGACTGCGCGAGAGGCCCAACTGA 687

RESULT 2
 ABL41173
 ID ABL41173 standard; cDNA; 687 BP.

XX ABL41173;
 AC 12-AUG-2002 (first entry)
 XX
 DE H. crista alternative fluorescent protein mutant FP10-cr1 cDNA.
 XX Stichodactyla; chromoprotein; fluorescent; anthozoan; food; FRET;
 KM colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
 KM fluorescence resonance energy transfer; gene expression; mutant; ss.
 XX
 OS Heteractis crista.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT /tag= a
 FT /product= "mutant fluorescent protein"
 FT misc_feature 4..6
 FT /tag= b
 FT /note= "the amino acid Val encoded by the above codon is
 FT not indicated in the corresponding protein"
 XX
 PN W0200230965-A2.
 XX
 PD 18-APR-2002.
 XX
 BP 12-OCT-2001; 2001MO-US032080.
 XX
 FR 12-OCT-2000; 2000US-024001BP.
 FR 16-JUL-2001; 2001US-0306131P.
 XX
 PA (CLON-) CLONTECH LAB INC.
 XX
 PI Lukyanov SA, Fradkov AF, Lukyanov KA, Gurekaya NG;
 DR WPI; 2002-444170/47.
 DR P-FSDB; ABB07995.
 XX
 PT Novel nucleic acid encoding Stichodactyla chromoprotein and its
 PT fluorescent mutant useful as coloring agent, labels in analyte detection
 PT assays, markers in recombinant DNA applications and filters in
 PT sunscreens.
 XX
 PS Claim 5; Fig 10; 81pp; English.
 XX
 CC The invention relates to a nucleic acid present in other than its natural
 CC environment and encoding an Stichodactyla chromoprotein or its
 CC fluorescent mutant, where the fluorescent protein has an emission maximum
 CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
 CC useful in applications employing a chromo or fluorescent nucleic acid or
 CC protein. Recombinant vectors comprising the nucleic acid is useful for
 CC producing an Anthozoan chromo and/or fluorescent protein. The
 CC chromoproteins, and their fluorescent mutants are useful as colouring
 CC agents capable of imparting colour or pigment to a particular composition
 CC of matter. The chromoproteins can be incorporated into a variety of
 CC different compositions including food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g. animals and plants, and as labels in
 CC analyte detection assays, e.g. assays for biological analyses of interest
 CC (see AB41167 for a detailed description of the various uses of the
 CC chromoproteins). The present sequence represents the H. crista
 CC alternative fluorescent protein mutant FP10-cr1 encoding cDNA
 CC
 SQ Sequence 687 BP; 145 A; 230 C; 219 G; 93 T; 0 U; 0 Other;

Query Match 97.7%; Score 671; DB 6; Length 687;
 Best Local Similarity 98.5%; Pred. No. 6.5e-109;
 Matches 677; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGTGAGCGGCGCTGTGAAGAGAGATGCGCATCAAGATGATCATGAGAGGACCGT 60
 DB 1 ATGTGAGCGGCGCTGTGAAGAGAGAGATGCGCATCAAGATGATCATGAGAGGACCGT 60
 QY 61 AACGGCCACTACTTCAAGTGCAGAGGCGGAGACCGCAACCCCTTGCCTGGACCCGAC 120

```
Db 61 AACGGCCACATCTTCAAGTCCGAGGCGGAGGCGGCAACCCCTTCGCGGACACCGAG 120
Qy 121 AGCATGAGAAATCCACGATGACCGAGGGGGCCCCCTGCGCTTCGATCGATCTGGGCG 180
Db 121 AGCATGCGAATCCAGTGAACGAGGGGCCCCCTTCGCTTCGATCGATCTGGGCG 180
Qy 181 CCTGCTGCGAGTACCGGACGAGACCTTCGTGCAACCAACCGCCGAGATCCCGCACTTC 240
Db 181 CCTGCTGCGAGTACCGGACGAGACCTTCGTGCAACCAACCGCCGAGATCCCGCACTTC 240
Qy 241 TTCAAGCAGAGCTTCCCGAGGGCTTCACCTGGAGAGAAACCACTTACGAGAGCGC 300
Db 241 TTCAAGCAGAGCTTCCCGAGGGCTTCACCTGGAGAGAAACCACTTACGAGAGCGC 300
Qy 301 GGCATCTGACCGGCCACGAGACCAACGACCTGAGAGGCAACGCTGATCTCAAGGTG 360
Db 301 GGCATCTGACCGGCCACGAGACCAACGACCTGAGAGGCAACGCTGATCTCAAGGTG 360
Qy 361 AAGGTGACCGGACCAACTTCCCGCGGACGCGCCCGTGTGAGAAACAAGACGCGCGC 420
Db 361 AAGGTGCTGGGACCAACTTCCCGCGGACGCGCCCGTGTGAGAAACAAGACGCGCGC 420
Qy 421 TGGGAGCCGACGACCGAGGTGTGTACCCCGAAGACGCGCTGTGCGCGCGAAGCTG 480
Db 421 TGGGAGCCGACGACCGAGGTGTGTACCCCGAAGACGCGCTGTGCGCGCGAAGCTG 480
Qy 481 ATGGCCCTGAAGGTGGGCGACCGGCGACCTGATCGGACCACTACCAAGCTACCGGAGC 540
Db 481 ATGGCCCTGAAGGTGGGCGACCGGCGCTATCTGCAACCACTACCAAGCTACCGGAGC 540
Qy 541 AAGAAGGCCCTGCGCGCCCTGACCATGCGCGGCTTCCACTTCAACCGACATCCGGCTCAG 600
Db 541 AAGAAGGCCCTGCGCGCCCTGACCATGCGCGGCTTCCACTTCAACCGACATCCGGCTCAG 600
Qy 601 ATCTGCGGAGAAAGAAAGACGAGTACTTGCAGCTGTACGAGGCCAGCGTGCCTCGGTAC 660
Db 601 ATCTGCGGAGAAAGAAAGACGAGTACTTGCAGCTGTACGAGGCCAGCGTGCCTCGGTAC 660
Qy 661 AGGACCTGCGCGGAGAAAGGCCCACTGA 687
Db 661 AGGACCTGCGCGGAGAAAGGCCCACTGA 687

RESULT 3
ABL41174
ID ABL41174 standard; DNA; 1396 BP.
XX
XX ABL41174;
XX
DT 12-AUG-2002 (first entry)
XX
DE Cr-449-tandem fusion protein nucleotide sequence.
XX
XX Stichodactyliden; chromoprotein; fluorescent; anthozoan; food; FRET;
KM colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
KM fluorescence resonance energy transfer; fusion protein; ds.
XX
XX Synthetic.
OS Heteractis crispa.
XX
XX
FH Key 14. 1396 Location/Qualifiers
FT CDS /*tag= a
FT misc_feature /product= "Cr-449-tandem fusion protein"
FT /tag= b
FT /note= "linker nucleotide fragment"
XX
XX W0200230965-A2.
XX
XX 18-APR-2002.
XX
```

```
PF 12-OCT-2001; 2001MO-US032080.
XX
XX 12-OCT-2000; 2000US-0240018P.
PR 16-JUL-2001; 2001US-0306131P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukanov SA, Fradkov AF, Lukanov KA, Gurskaya NG;
PI WPI; 2002-444170/47.
XX P-PSDB; ABB07996.
XX
XX Novel nucleic acid encoding Stichodactyliden chromoprotein and its
PT fluorescent mutant useful as coloring agent, labels in analyte detection
PT assays, markers in recombinant DNA applications and filters in
PT sunscreens.
XX
XX Disclosure; Fig 12; 81pp; English.
XX
XX The invention relates to a nucleic acid present in other than its natural
CC environment and encoding an Stichodactyliden chromoprotein or its
CC fluorescent mutant, where the fluorescent protein has an emission maximum
CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
CC useful in applications employing a chromo or fluorescent nucleic acid or
CC protein. Recombinant vectors comprising the nucleic acid is useful for
CC producing an Anthozoan chromo and/or fluorescent protein. The
CC chromoproteins, and their fluorescent mutants are useful as coloring
CC agents capable of imparting colour or pigment to a particular composition
CC of matter. The chromoproteins can be incorporated into a variety of
CC different compositions including food compositions, pharmaceuticals, in
CC cosmetics, living organisms, e.g. animals and plants, and as labels in
CC analytical detection assays, e.g. assays for biological analyses of interest
CC (see ABL41167 for a detailed description of the various uses of the
CC chromoproteins). The present sequence represents the H. crispa Cr-449-
CC tandem fusion protein nucleotide sequence
XX
XX Sequence 1396 BP; 293 A; 471 C; 444 G; 188 T; 0 U; 0 Other;
SQ
Query Match 97.7%; Score 671; DB 6; Length 1396;
Best Local Similarity 98.5%; Pred. No. 6,5e-109;
Matches 677; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 1 ATGTGTAGGCGGCTGTGAAGAGAGATATGCGCATCAGATGTATCATGAGGGGCAACCGTG 60
Db 710 ATGTGTAGGCGGCTGTGAAGAGAGATATGCGCATCAGATGTATCATGAGGGGCAACCGTG 769
Qy 61 AACGGCCACTACTTCAAGTGGCGAGGGCGGAGGCGGCAACCCCTTCCCGGCAACCGG 120
Db 770 AACGGCCACTACTTCAAGTGGCGAGGGCGGAGGCGGCAACCCCTTCCCGGCAACCGG 829
Qy 121 AGCATGAGAAATCCAGTGAACGAGGGGCGCCCGCTGCGCTTCGACATCTCGAGC 180
Db 830 AGCATGCGAATCCAGTGAACGAGGGGCGCCCGCTGCGCTTCGACATCTCGAGC 889
Qy 181 CCTGCTGCGAGTACCGGACGAGACCTTCGTGCAACCAACCGCCGAGATCCCGCACTTC 240
Db 890 CCTGCTGCGAGTACCGGACGAGACCTTCGTGCAACCAACCGCCGAGATCCCGCACTTC 240
Qy 241 TTCAAGCAGAGCTTCCCGAGGGCTTCACCTGGAGAGAAACCACTTACGAGAGCGC 300
Db 950 TTCAAGCAGAGCTTCCCGAGGGCTTCACCTGGAGAGAAACCACTTACGAGAGCGC 300
Qy 301 GGCATCTGACCGGCCACGAGACCAACGACCTGAGAGGCAACGCTGATCTCAAGGTG 360
Db 1010 GGCATCTGACCGGCCACGAGACCAACGACCTGAGAGGCAACGCTGATCTCAAGGTG 1069
Qy 361 AAGGTGACCGGACCAACTTCCCGCGGACGCGCCCGTGTGAGAAACAAGACGCGCGC 420
Db 1070 AAGGTGCTGGGACCAACTTCCCGCGGACGCGCCCGTGTGAGAAACAAGACGCGCGC 1129
Qy 421 TGGGAGCCGACGACCGAGGTGTGTACCCCGAAGACGCGCTGTGCGCGCGAAGCTG 480
Db 1130 TGGGAGCCGACGACCGAGGTGTGTACCCCGAAGACGCGCTGTGCGCGCGAAGCTG 1189
```

```

QY 481 ATGGCCCTGAAAGGTGGGCGACCGGACCTGATCTGCCACCACTACACGAGTACCGGAGC 540
DB 1190 ATGGCCCTGAAAGGTGGGCGACCGGCGGTGATCTGCCACCACTACACGAGTACCGGAGC 1249
QY 541 AAGAAAGCCGCTGGCGCCGCTGACCAATGCCCGGCTTCCACTTCAACGACATCCGGGTCCAG 600
DB 1250 AAGAAAGCCGCTGGCGCCGCTGACCAATGCCCGGCTTCCACTTCAACGACATCCGGGTCCAG 1309
QY 601 ATGCTGGGAAAGAAAGAGACGATCTTGCAGCTGTACGAGGCCAGCGGTGGCCCGGTAC 660
DB 1310 ATGCTGGGAAAGAAAGAGACGATCTTGCAGCTGTACGAGGCCAGCGGTGGCCCGGTAC 1369
QY 661 AGCGACCTGGCCGAGAAAGGCCCAACTGA 687
DB 1370 AGCGACCTGGCCGAGAAAGGCCCAACTGA 1396

```

RESULT 4

ACA61024 standard; DNA; 1396 BP.

ACA61024;

09-JUL-2003 (first entry)

DNA encoding chromoprotein Cr-449-tandem.

Chromo/fluorescent domain; labeled fusion protein;
site-specific gene modification; chromoprotein; colouring agent;
food composition; pharmaceutical; cosmetic; Cr-449-tandem; gene; ds.
Anthozoa.

Location/Qualifiers

Key

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

WT CDS

selectable markers in recombinant DNA applications. This sequence encodes
chromo/fluorescent domain fusion protein Cr-449-tandem

Sequence 1396 BP; 293 A; 471 C; 444 G; 188 T; 0 U; 0 Other;

Query Match 97.7%; Score 671; DB 7; Length 1396;
Best Local Similarity 98.5%; Pred. No. 6,5e-109;
Matches 677; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

QY 1 ATGTGAGCGGCGCTGCTGAAGAGAGATGTCGATCAATGATATGATGAGAGGACCGTG 60
DB 710 ATGTGAGCGGCGCTGCTGAAGAGAGATGTCGATCAATGATATGATGAGAGGACCGTG 769
QY 61 AAGCGCACTACTTCAAGTGGAGGCGAGGCGGACAGGACCCCTTGGCCGACCCAG 120
DB 770 AAGCGCACTACTTCAAGTGGAGGCGAGGCGGACAGGACCCCTTGGCCGACCCAG 829
QY 121 AGCATGAGATTCACGTCGACCGAGGCGCCCTTGGCCCTTGGCCCTTGGCATTCGGCC 180
DB 830 AGCATGAGATTCACGTCGACCGAGGCGCCCTTGGCCCTTGGCCCTTGGCATTCGGCC 889
QY 181 CCTGCTCGAGTACGCGCAGCAGGACCTTGTGACACACACGCGAGATCCCGACTTC 240
DB 890 CCTGCTCGAGTACGCGCAGCAGGACCTTGTGACACACACGCGAGATCCCGACTTC 949
QY 241 TTCAAGCAGAGCTTCCCGAGGCGCTTCACTGAGAGAGAACACCACTTACGAGAGCGC 300
DB 950 TTCAAGCAGAGCTTCCCGAGGCGCTTCACTGAGAGAGAACACCACTTACGAGAGCGC 1009
QY 301 GGATATCTGAGACCGGCGCAGGACACAGACCTGAGAGGCGCACTGCTGATCTTCAAGGTG 360
DB 1010 GGATATCTGAGACCGGCGCAGGACACAGACCTGAGAGGCGCACTGCTGATCTTCAAGGTG 1069
QY 361 AAGGTGACGCGGACCACTTCCCGCGAGCGGCGCCCTGATGAAAGAACAGAGCGCGCG 420
DB 1070 AAGGTGACGCGGACCACTTCCCGCGAGCGGCGCCCTGATGAAAGAACAGAGCGCGCG 1129
QY 421 TGGAGGCCAGCAGCAGGAGTGTGTACCCCGAGACGCGGTGTGTGTGTGTGTGTGTGTGT 480
DB 1130 TGGAGGCCAGCAGCAGGAGTGTGTACCCCGAGACGCGGTGTGTGTGTGTGTGTGTGT 1189
QY 481 ATGGCCCTGAAAGGTGGGCGACCGGACCTGATCTGCCACCACTACACGAGTACCGGAGC 540
DB 1190 ATGGCCCTGAAAGGTGGGCGACCGGCGGTGATCTGCCACCACTACACGAGTACCGGAGC 1249
QY 541 AAGAAAGCCGCTGGCGCCGCTGACCAATGCCCGGCTTCCACTTCAACGACATCCGGGTCCAG 600
DB 1250 AAGAAAGCCGCTGGCGCCGCTGACCAATGCCCGGCTTCCACTTCAACGACATCCGGGTCCAG 1309
QY 601 ATGCTGGGAAAGAAAGAGACGATCTTGCAGCTGTACGAGGCCAGCGGTGGCCCGGTAC 660
DB 1310 ATGCTGGGAAAGAAAGAGACGATCTTGCAGCTGTACGAGGCCAGCGGTGGCCCGGTAC 1369
QY 661 AGCGACCTGGCCGAGAAAGGCCCAACTGA 687
DB 1370 AGCGACCTGGCCGAGAAAGGCCCAACTGA 1396

```

RESULT 5

ABL41175 standard; DNA; 1424 BP.

ABL41175;

12-AUG-2002 (first entry)

Cr-449-tandem-actin fusion protein nucleotide sequence.

Stichodactyliden; chromoprotein; fluorescent; anthozoa; food; FRRT;
colouring agent; pharmaceutical; cosmetic; immunosay; biosensor; gene;
fluorescence resonance energy transfer; fusion protein; ds.
Synthetic.

XX Lutyakov SA;
PI
XX WPI; 2003-381709/36.
DR P-PSDB; ABU09922.
XX
XX
PT New nucleic acid encoding polypeptide products having at least two linked
PT chromo/fluorescent domains, useful for generating transgenic plants or
PT animals or site-specific gene modifications in cell lines.
XX
PS Disclosure; Fig 2; 68pp; English.

CC The invention describes a nucleic acid encoding a polypeptide product
CC comprising a first and a second chromo/fluorescent domain, optionally
CC joined by a linking domain. The first and second chromo/fluorescent
CC domains associate with each other under intracellular conditions so that
CC the encoded polypeptide assumes a tertiary structure. The nucleic acid
CC and the protein are useful in producing labeled fusion proteins that have
CC a precise and predictable signal to fusion partner ratio. The nucleic
CC acid may also be used in generating transgenic, non-human plants or
CC animals or site-specific gene modifications in cell lines. The
CC chromoproteins may be used as colouring agents, as a food composition, in
CC pharmaceuticals or cosmetics, as labels in analyte detection assays or as
CC selectable markers in recombinant DNA applications. This sequence encodes
CC chromo/fluorescent domain fusion protein Cf-449-tandem-actin
XX
XX
50 Sequence 1424 BP; 301 A; 477 C; 452 G; 194 T; 0 U; 0 Other;

Query Match	97.4%	Score 669.4	DB 7	Length 1424
Best Local Similarity	98.4%	Pred. No. 1.2e-108		
Matches 676; Conservative	0	Mismatches 11	Indels 0	Gaps 0

QY	1	TTGTGAGCGCGCTGCTGAAGAGAGATGTGGCATCAATGATATATGAGAGGACACCTG	60
Db	14	ATGTGTAGCGCGCTGCTGAAGAGAGATGTGGCATCAATGATATATGAGAGGACACCTG	73
QY	61	AAGGSCACATCTTCAATGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	120
Db	74	AAGGSCACATCTTCAATGTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	133
QY	121	AGCTTGAAGATCCAGTGAACCGAGAGGCGCCCCCTGCGCTTGCCTTGAATCTTGAGCC	180
Db	134	AGCATGTGGATCCAGTGAACCGAGAGGCGCCCCCTGCGCTTGCCTTGAATCTTGAGCC	193
QY	181	CCCTGTGCGAGTACCGGACGAGACCTTGTGTGACACACCGCGAGAACTCCGACCTTC	240
Db	194	CCCTGTGCGAGTACCGGACGAGACCTTGTGTGACACACCGCGAGAACTCCGACCTTC	253
QY	241	TTCAAGACAGAGCTCTCCCGAGGGCTTCACTGTGGAGAGAACCAACCACTTACGAGACGCG	300
Db	254	TTCAAGACAGAGCTCTCCCGAGGGCTTCACTGTGGAGAGAACCAACCACTTACGAGACGCG	313
QY	301	GGCATCTGTACCGGCCACCAAGAACCAAGCTGTGAGAGGCAACTGCTGTATCTTCAAGAGTG	360
Db	314	GGCATCTGTACCGGCCACCAAGAACCAAGCTGTGAGAGGCAACTGCTGTATCTTCAAGAGTG	373
QY	361	AAAGTCACAGGCAACCACTTCCCGCGACGCGCCCGGTATGAAGAACAAAGACGCGCGC	420
Db	374	AAAGTCGTGGGACCACTTCCCGCGAGGCGCCCGGTATGAAGAACAAAGACGCGCGC	433
QY	421	TGGAGCCCAAGACCGAGGTGTATCCCGAGAACCGCGTGTGTGCGCGCGGAACGTG	480
Db	434	TGGAGCCCAAGACCGAGGTGTATCCCGAGAACCGCGTGTGTGCGCGCGGAACGTG	493
QY	481	ATGGCCCTGAAGTGGGGAACCGGACCTGTATCTGACACCTTACACAGGTATCCGAGGC	540
Db	494	ATGGCCCTGAAGTGGGGAACCGGCGGCTGTATCTTCCACACCTTACACAGGTATCCGAGGC	553
QY	541	AAAGAGGCGGTGCGGCGCTGTACCAATGCCGCTTCCACTTCAACCGACATCCGAGTCCAG	600
Db	554	AAAGAGGCGGTGCGGCGCTGTACCAATGCCGCTTCCACTTCAACCGACATCCGAGTCCAG	613
QY	601	ATGCTGCGGAGAAAGAGACGAGTACTTTCGAGCTGTATCGAGGCGACGCTGGCCCGGATAC	660

Db 614 ATGCTGCGGAGGAGGAGGACGAGTACTTCAGCTGTACGAGGCCAGCTGCGCGGTAC 673

Qy 661 AGCGACTGTGCCCCGAGGAGGCCCAACTGA 687

Db 674 AGCGACTGTGCCCCGAGGAGGCCCAACGA 700

RESULT 7
ACA61026
ID ACA61026 standard; DNA; 1376 BP
yy

DT 09-JUL-2003 (first entry)
XX

DE DNA encoding chromoprotein HcRed-Cr1-tandem
XX

Chromo/fluorescent domain; labeled fusion protein;
site-specific gene modification; chromoprotein; colouring agent;
food composition; pharmaceutical; cosmetic; HcRed-Cr1-tandem; gene; ds

OS Anthozoa.

	Key	Location/Qualifiers
FH		
FT		
FT	CDS	1. .1376
		/+tag= a

```

FT      /product= "HcRed-Cr1-tandem"
FT      /transl_except= (pos:679, .680, aa:Asn)
XX

```

PN W02003031590-A2.

PD 17-APR-2003

PF 10-OCT-2002; 2002WO-US032560.
XX

PR 12-OCT-2001; 2001US-00976673.
DB 11-FEB-2003; 2003US-03553355

PR 22-MAY-2002; 2002US-0383336P.
XY

PA (CLON-) CLONTech LAB INC.
XY

PI Lukyanov SA;
XY

DR WPI; 2003-381709/36
DR R-PCDB; 2003-381709/36

XX
BT
New York

chromo/fluorescent domains, useful for generating transgenic plants or animals or into specific gene constructs.

PS Disclosure; Fig 3; 68pp; English.
yy

The invention describes a nucleic acid encoding a polypeptide product comprising a first and a second chromo/fluorescent domain, optionally joined by a linking domain. The first and second chromo/fluorescent domains associate with each other under intracellular conditions so that the encoded polypeptide assumes a tertiary structure. The nucleic acid and the protein are useful in producing labeled fusion proteins that have a precise and predictable signal to fusion partner ratio. The nucleic acid may also be used in generating transgenic, non-human plants or animals or site-specific gene modifications in cell lines. The chromoproteins may be used as colouring agents, as a food composition, in pharmaceuticals or cosmetics, as labels in analyte detection assays or as selectable markers in recombinant DNA applications. This sequence encodes chromo/fluorescent domain fusion protein fKced-Cy1-tandem

Sequence	1376 BP, 381 A, 292 C, 357 G, 346 T, 0 U, 0 Other;
Query Match	55.0%; Score 446.6; DB 7; Length 1376;
Best Local Similarity	Pred. No. 146-69; 78.2%;
Matches	536; Conservative 149; Indels 0; Gaps 0;

QY 3 GGTGACGGCTCTCTGAGAGATATGCCATCAAGATGTATATGAGGGCCACCTGAA 62
 DB 692 GATGCTGTGTTGTGTGAGAGAGATATGCGCATCAAGATGTATATGAGGGCCACCTGAA 751
 QY 63 CGGCACTTCAAGTGGAGGGGCGAGGGGCGAGGCAACCTTGGCGGGCGAGAG 122
 DB 752 TGGCCATTATTTTATGAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 811
 QY 123 CATGAGATTCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182
 DB 812 CATGAGATTCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
 QY 183 CTGCTGCAATGAG 242
 DB 872 GTGCTTGTGAGTGTAG 931
 QY 243 CAGGAG 302
 DB 932 CAGGAG 991
 QY 303 CATCTGACCGCCCGAG 362
 DB 992 CATCTTATCTGCTCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1051
 QY 363 GGTGACGAG 422
 DB 1052 AGTCATGTGATCAATTTTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1111
 QY 423 GAG 482
 DB 1112 GAG 1171
 QY 483 GGGCTGAG 542
 DB 1172 GGGCTTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1231
 QY 543 GAG 602
 DB 1232 GAG 1291
 QY 603 GGTGCGGAG 662
 DB 1292 GGTGAG 1351
 QY 663 CGAGCTGCGGAG 687
 DB 1352 TGATCTTCTGAG 1376
 RESULT 8
 ABL41170
 ID ABL41170 standard; cDNA; 681 BP.
 AC ABL41170;
 XX
 DT 12-AUG-2002 (first entry)
 XX
 DE H. crispata fluorescent protein mutant 44-9 encoding cDNA.
 XX
 KW Stichodactyliden; chromoprotein; fluorescent; anthozoan; food; FRET;
 KM colouring agent; pharmaceutical; cosmetic; immunosensor; biosensor; gene;
 KM fluorescence resonance energy transfer; gene expression; mutant; ss.
 XX
 OS Heteractis crispata.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..681
 FT /*tag= a
 FT /product= "mutant fluorescent protein"
 XX
 PN MO200230965-A2.
 XX
 PD 18-APR-2002.

XX 12-OCT-2001; 2001WO-US032080.
 PF 12-OCT-2001; 2000US-024018P.
 XX 16-JUL-2001; 2001US-0306131P.
 PR (CLON-) CLONTECH LAB INC.
 XX
 PA LUKYANOV SA, FRADKOV AF, LUKYANOV KA, GURSKAYA NG;
 PI MPI; 2002-444170/47.
 DR P-PSDB; ABB07992.
 XX
 PT Novel nucleic acid encoding Stichodactyliden chromoprotein and its
 PT fluorescent mutant useful as coloring agent, labels in analyte detection
 PT assays, markers in recombinant DNA applications and filters in
 PT sunscreens.
 XX
 PS Claim 5; Fig 6; 8hpp; English.
 XX
 CC The invention relates to a nucleic acid present in other than its natural
 CC environment and encoding an Stichodactyliden chromoprotein or its
 CC fluorescent mutant, where the fluorescent protein has an emission maximum
 CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
 CC useful in applications employing a chromo or fluorescent nucleic acid or
 CC protein. Recombinant vectors comprising the nucleic acid is useful for
 CC producing an Anthozoan chromo and/or fluorescent protein. The
 CC chromoproteins, and their fluorescent mutants are useful as colouring
 CC agents capable of imparting colour or pigment to a particular composition
 CC of matter. The chromoproteins can be incorporated into a variety of
 CC cosmetic compositions including food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g. animals and plants, and as labels in
 CC analytical detection assays, e.g. assays for biological analytes of interest
 CC (see ABL41167 for a detailed description of the various uses of the
 CC chromoproteins). The present sequence represents the H. crispata
 CC fluorescent protein mutant 44-9 encoding cDNA
 XX
 SQ Sequence 681 BP; 186 A; 144 C; 178 G; 173 T; 0 U; 0 Other;
 Query Match 64.2%; Score 441.2; DB 6; Length 681;
 Best Local Similarity 78.2%; Pred. No. 1.3e-68;
 Matches 530; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
 QY 10 GGCCTGCTGAGAGAGAGATGCGCATCAAGATGTATATGAGGGGCAACCTGAGACGGCCAC 69
 DB 4 GGTGTTGTAAGAGAGATGCGCATCAAGATGTATATGAGGGGCAACCTGATATGCCAT 63
 QY 70 TACTTCAAGTGTGAG 129
 DB 64 TATTTCAAGTGTGAG 123
 QY 130 ATCCAGGTGAG 189
 DB 124 ATTCATGTACCGAG 183
 QY 190 GAGTACGAG 249
 DB 184 GAGTACGAG 243
 QY 250 AGCTTCCCGAG 309
 DB 244 TCTTTCCCGAG 303
 QY 310 ACCGCCACGAG 369
 DB 304 ACTGCTCATCAG 363
 QY 370 GGCACCAACTTCCCGAG 429
 DB 364 GGTACCAATTTTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
 QY 430 AGCACGAG 489

RESULT 9
ABL41171
ID ABL41171 standard; cDNA; 681 BP.

Stichodactylidae; chromoprotein; fluorescent; anthozoan; food, FRET; colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene fluorescence resonance energy transfer; gene expression; mutant; ss.

Key	Location/Qualifiers
FH	1..681
FT	/*tag= a
FT	/product= "mutant fluorescent protein"
FT	

PA (CLON-) CLONTECH LAB INC.

DR WPI; 2002-444170/47.
DR P-PSDB; ABB07993.

PT Novel nucleic acid encoding Stichodactyla den chromoprotein and its fluorescent mutant useful as coloring agent, labels in analyte detection PT assays, markers in recombinant DNA applications and filters in PT sunscreens.

PS Claim 5; Fig 8; 81pp; English.

The invention relates to a nucleic acid present in other than its natural environment and encoding an Stichodactyliden chromoprotein or its fluorescent mutant, where the fluorescent protein has an emission maximum ranging from 580-660 nm. The polynucleotides and encoded proteins are useful in applications employing a chromo or fluorescent nucleic acid or protein. Recombinant vectors comprising the nucleic acid is useful for producing an Anthozoan chromo and/or fluorescent protein. The chromoproteins, and their fluorescent mutants are useful as colouring agents capable of imparting colour or pigment to a particular composition of matter. The chromoproteins can be incorporated into a variety of

Query Match	64.0%;	Score 439.6;	DB 6;	Length 681;
Best Local Similarity	78.0%;	Pred. No. 2.4e-68;		
Matches 529; Conservative	0;	Mismatches 149;	Indels 0;	Gaps 0

QY 10 GGCCTGCGAAGGAGCTGTGGCATCAATATTAATAGAGGACCGTGAAACGGGCAC 69
Db 4 GGGTTTGTGAAAGAAAGTATGGCGATCAAGATGTACATGGAAAGGACCGATTATGGCGCAT 63
QY 70 TACTTCAAGTGCAGAGGGCGAAGGGCGACCGCATCCCTTGGCCGCGACCCAGAGCATAGAG 129
Db 64 TATTTTCAAGTGTAAAGAAAGGAGAACCGGACCACTTGGCAGGATACCAAGACATAGAGG 123
QY 130 ATTCACGTGACCGAAGGGCGCCCCCTGTGCGCTTGGCTTGAACATCTTGAGCCCCCTGTGCG 189

Db 124 ATTCATGTCACCGAAGG

190 GAGTACGGCAGCAGGAC

Db 184 GCGTACGGCAGCAGGAC

250

244

210

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

7
8
9
0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9
0

[illegible]

DD 424 AGCACIGAGGIGGIIYA

Qy 490 AAGTGGGCGACCGGCA

D5 484 AAAGTCGGTGATCGTCC

QY 550 GTGCGCGCCCTGACCAT

Db 544 GTCCGTGCCTTGACAAAT

QY 610 AAGAAGAAGCAGCTA

604 AAAGAGAAAGACGAGTA

670 CCGGAGGAGGCCAATTG

664

RESULT 10

ID . ABL41180 standard; cDNA; 6

AC ABL41180;

DT 12-AUG-2002 (first entry)

XX
DE H
cristina chromoprotein mi

XX Chichobutun: shomen

KM	colouring agent; pharmaceutical; cosmetic; immunoassay; bioensor; gene;
KW	fluorescence resonance energy transfer; gene expression; mutant; ss.
XX	
OS	Heteractis crispus.
XX	
FH	Key
FT	Location/Qualifiers
CDS	1..684
FT	/*tag= a
PT	/product= "mutant fluorescent protein"
XX	
PN	WO200230965-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US032080.
XX	
PR	12-OCT-2000; 2000US-024001BP.
PA	16-JUL-2001; 2001US-030613IP.
XX	
XX	(CLON-) CLONTECH LAB INC.
PI	Lukyanov SA, Pradkov AF, Lukyanov KA, Gurskaya NG;
XX	
DR	WPI; 2002-444170/47.
DR	P-PDB; ABB07998.
XX	
XX	Novel nucleic acid encoding Stichodactylidaeen chromoprotein and its
PT	fluorescent mutant useful as coloring agent, labels in analyte detection
PT	assays, markers in recombinant DNA applications and filters in
PT	screens.

PS Example; fig 15; 81bp; English.
 CC
 CC The invention relates to a nucleic acid present in other than its natural
 CC environment and encoding an Stichodactylaaden chromoprotein or its
 CC fluorescent mutant, where the fluorescent protein has an emission maximum
 CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
 CC useful in applications employing a chromo or fluorescent nucleic acid or
 CC protein. Recombinant vectors comprising the nucleic acid is useful for
 CC producing an Anthozoan chromo and/or fluorescent protein. The
 CC chromoproteins, and their fluorescent mutants are useful as colouring
 CC agents capable of imparting colour or pigment to a particular composition
 CC of matter. The chromoproteins can be incorporated into a variety of
 CC different compositions including food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g. animals and plants, and as labels in
 CC analytical detection assays, e.g. assays for biological analytes of interest
 CC (see ABL41167 for a detailed description of the various uses of the
 CC chromoproteins). The present sequence represents the cDNA encoding an
 CC alternative embodiment of the H. cripsa fluorescent protein mutant C148S
 CC
 QQ Sequence 684 BP; 188 A; 145 C; 178 G; 173 T; 0 U; 0 Other;

	Query Match	53.8%	Score 438.2	DB 6	Length 684
	Best Local Similarity	77.6%	Fid. No. 4.3e-68		
	Matches 530	Conservative 0	Mismatches 153	Indels 0	Gaps 0
Qy	5	TGACCGGCGCTCTGAGAGAGATAGCCCATCAAGATGTACATGAGGGCACCTGTAAAC	64		
Db	2	TGGCTGGTTTGTGAAAGAAAGTATGGCCATCAAGTATACATGAAAGGACACGGTTAATG	61		
Qy	65	GCCACTACTTCAATGTCGACGGGGCCGACGGCAACCCCTTGGCCCGGACCCGAGACA	124		
Db	62	GCCATTATTTAAAGTGTAAGAGAGAGGAGACGGGACCCCATTTTACAGATGACGACAGACA	121		
Qy	125	TGAGAATTCACGTGACCGAGGGGCGCCCTCTGCCTTCGCTTCGACATCTCTGAGCCGCT	184		
Db	122	TGAGGATTCATGTGACCGAAGGGGCTCATTAACATTTGCTTCGACATTTTGGACCGT	181		
Qy	185	GCTCGCAATAGGCGACGACGACCTTCGTGCACACACCGCGGAGATCCCGCACTTCTTCA	244		
Db	182	GTTTGATTAACGGCGACGACCTTTGTCTCCACCAATAGGACAGAGATTTCCGATTTTCTTCA	241		
Qy	245	AGCAGAGCTTCCCGAGGGGCTTCACTCTGGAGAGAACCAACCACTTACGAGACGCGGCA	304		

DB 242 AGCAGTCTTTCCCTGGAAGCTTTACTTTGGGAAAGAACCAACTATGAAAGATGAGGCA 301

OY 305 TCCTGACCGGCCACAGGACACACAGCCTGGAGGGCAACTGCGTGAATCTTCAAGAGTGAAAG 364

DB 302 TTTCTTACTGCTCATAGGACACACAGCCTGGAGGGAACTGCTTTATATCAAGGTGAAAG 361

OY 365 TGCAAGGACCAACTTCCCCGCGCAAGCGCCCGTGATGAAAGACAAGACGGCGGCTGGG 424

DB 362 TCCTTGTATCCAAATTTTCTGTGTATGGCCCGTGATGAAAGAAACAATACGAGGAGATGGG 421

OY 425 AGCCAGACACGAGGTGGTGTATACCCCGAAGACGGCGTGTGTGCGGCCGGAACGTATGG 484

DB 422 AGCCAGACACTGAGGTGGTGTATCCAGAGAAATGGTGTCTGTGTGACCTTAATGTATGG 481

OY 485 CCTGGAAGTGGGCGACCGGCACTGTATCTGCAACACATCACACCACTAACCGGAGGAAG 544

DB 482 CCTTTAAAGTCGTATCGTGTGTATCTGTGCATCTTATCTTCTTACAGGTCCAAAG 541

OY 545 AGGCGGTGCGCGCCCTGACCAATGCCCGGCTTCACATTCACGCAATCCGAGTCCAGATGC 604

DB 542 AAGCAGTCCGTGCTTGAACAATGCCAGGATTTTATCAAGACATCCGCTTCAAGTGC 601

OY 605 TGCGGAAGAAGAGACGATCTTGAAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCG 664

DB 602 CGAGGAAAAAGAAACAGAGTACTTTGAACTGTACGAAGCACTGTGTGGTACAGTG 661

OY 665 AACTGCGCGAAGGCCCACTGA 687

DB 662 ATCTTCTGAAAAAGCAATGA 684

RESULT 11	
ABL41169	
ID	ABL41169 standard; cDNA; 684 BP.
XX	
AC	ABL41169;
XX	
DT	12-AUG-2002 (first entry)
XX	
DE	H. crista fluorescent protein mutant C148S encoding cDNA.
XX	
KW	Stichodactylidae; chromoprotein; fluorescent; anthozoan; food; FRET;
KW	colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
KW	fluorescence resonance energy transfer; gene expression; mutant; ss.
XX	
OS	Heteractis crispa.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	1..684
FT	/*tag= a
FT	/product= "mutant fluorescent protein"
XX	
PN	WO200230965-A2.
XX	
PD	18-APR-2002.
XX	
PF	12-OCT-2001; 2001WO-US032080.
XX	
PR	12-OCT-2000; 2000US-0240018P.
PR	16-JUL-2001; 2001US-0306131P.
XX	
PA	(CLON-) CLONTECH LAB INC.
XX	
PI	Lukyanov SA, Fradkov AF, Lukyanov KA, Gurskaya NG;
XX	
DR	WPI; 2002-444170/47.
XX	
DR	P-PSDB; ABB07991.
XX	
PT	Novel nucleic acid encoding Stichodactylidae chromoprotein and its
PT	fluorescent mutant useful as coloring agent, labels in analyte detection
PT	assays, markers in recombinant DNA applications and filters in
PT	screens.

```
XX Claim 5, Fig 4, 81pp; English.
PS
CC The invention relates to a nucleic acid present in other than its natural
CC environment and encoding an Stichodactylidae chromoprotein or its
CC fluorescent mutant, where the fluorescent protein has an emission maximum
CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
CC useful in applications employing a chromo or fluorescent nucleic acid or
CC protein. Recombinant vectors comprising the nucleic acid is useful for
CC producing an Anthozoan chromo and/or fluorescent protein. The
CC chromoproteins, and their fluorescent mutants are useful as colouring
CC agents capable of imparting colour or pigment to a particular composition
CC of matter. The chromoproteins can be incorporated into a variety of
CC different compositions including food compositions, pharmaceuticals,
CC cosmetics, living organisms, e.g. animals and plants, and as labels in
CC analyte detection assays, e.g. assays for biological analytes of interest
CC (see AB41167 for a detailed description of the various uses of the
CC chromoproteins). The present sequence represents the H. crista
CC fluorescent protein mutant C1485 encoding cDNA
XX
SQ Sequence 684 BP; 187 A; 146 C; 178 G; 173 T; 0 U; 0 Other;
Query Match 63.6%; Score 436.6; DB 6; Length 684;
Best Local Similarity 77.5%; Pred. No. 8.2e-68;
Matches 529; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 5 TGAGCGGCGCTGCTGAAGAGATGTGCGCATCAAGATGTATGAGAGGCGACCGTGAACG 64
DB 2 TGGCTGCTTTGTGTGAAGAAAGTATGCGCATCAAGATGTATGAGAGGCGACCGTTAATG 61
QY 65 GCCACTACTTCAAGTGCAGAGGCGAGGCGAACCCCTTGCCTGCGCACCCAGXGCA 124
DB 62 GCATTATTTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
QY 125 TGAGATTCACGTGATCCGAGAGGCGCCCGCTGCTTGCCTTGCATCTCTGCGCCCT 184
DB 122 TGAGATTCATGTACCGAGAGGCGCTCATTTACATTTGCGACATTTTGGCACCGG 181
QY 185 GCTGCGAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
DB 182 GTTGTAGATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
QY 245 AGCAGAGGCTCCCGAGAGGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
DB 242 AGCAGGCTTCCCTGAGAGGCTTACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
QY 305 TCCCTGACCGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
DB 302 TTTCTACTGCTCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361
QY 365 TGCACGAGACCACTTCCCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
DB 362 TCCCTGAGTCAATTTTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421
QY 425 AGCCAGAGACCGAGAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
DB 422 AGCAGAGACCTGAGAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
QY 485 CCTGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
DB 482 CCTTAAAGTCGATGATGCTGTTGATTCGCATCTTAACTTCTTAAAGTCCAGAGAG 541
QY 545 AGGCGGTGGCGCCCTGACATGCGCGGCTTCACTTCACTGAGAGAGAGAGAGAGAGAG 604
DB 542 AAGCAGTCCGTCCTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 601
QY 605 TGCAGAGAGAGAGAGAGAGATCTTCAAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAG 664
DB 602 CAGAGAGAGAGAGAGAGAGATCTTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
QY 665 ACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 687
DB 662 ATCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 684
```

```
RESULT 12
ABA00802
ID ABA00802 standard; cDNA; 760 BP.
AC ABA00802;
XX 01-APR-2003 (first entry)
DE Wild type chromoprotein cDNA.
XX Gene; kindling fluorescent protein; kindling stimulus; movement;
XX labeling; fluorescence resonance energy transfer; FRET;
XX bioluminescence resonance energy transfer; BRET; biosensor;
XX automated screening; ss.
XX Heteractis crispa.
XX
XX Key Location/Qualifiers
XX CDS 77..760
XX /tag= a
XX /product= "Chromoprotein"
XX
XX WO200296924-A1.
XX
XX 05-DEC-2002.
XX
XX 24-MAY-2002; 2002MO-US016379.
XX
XX 25-MAY-2001; 2001US-0293752P.
XX
XX 11-OCT-2001; 2001US-0329176P.
XX
XX (CLON-) CLONTECH LAB INC.
XX
XX Lukyanov SA, Chudakov D, Lukyanov K;
XX WPI; 2003-156788/15.
XX P-PSDB; AAG79765.
XX
XX Novel nucleic acid that is present in other than its natural environment
XX and that encodes kindling fluorescent protein, is useful in labeling
XX protocols, e.g. labeling proteins, organelles, cells and organisms.
XX
XX Example; Fig 4; 96pp; English.
XX
XX The sequences given in ABA00799-805 encode wild type and mutant kindling
XX fluorescent proteins. The proteins go from a first substantially non-
XX fluorescent or non-fluorescent state to a second fluorescent state upon
XX exposure to a kindling stimulus. The kindling proteins are useful for
XX detecting an entity such as a protein, organelle or cell in a composition
XX such as a cell or a multicellular composition (preferably a multicellular
XX organism), by providing the entity as an entity labeled with the kindling
XX protein, kindling the kindling fluorescent protein label with a kindling
XX stimulus to produce a kindling fluorescent protein label with a kindling
XX detecting the kindling fluorescent protein label with light and
XX monitoring the movement of the entity. The fluorescent proteins and the
XX cDNA encoding them are useful in labeling protocols, e.g., labeling
XX proteins, organelles, cells and organisms, as biological labels or
XX markers, in protein labeling or tagging applications. The fluorescent
XX kindling proteins are useful as detectable labels, as labels in analyte
XX detection assays, in fluorescence resonance energy transfer (FRET)
XX applications, as biosensors in prokaryotic and eukaryotic cells, in
XX applications involving the automated screening of arrays of cells
XX expressing fluorescent reporting groups, in high through-put screening
XX assays, as second messenger detectors, and in fluorescent activated cell
XX sorting assays
XX
XX Sequence 760 BP; 209 A; 166 C; 189 G; 196 T; 0 U; 0 Other;
Query Match 63.6%; Score 436.6; DB 7; Length 760;
```

	Best Local Similarity	77.5%	Pred. No. 8-2e-68;	
	Matches	529;	Conservative	0; Mismatches 154; Indels 0; Gaps 0;
QY	5	TGAGCGGCTCTCTGAGAGAGTATGCGCATCAAGATGTACATGGAGGGCAACCGTGAACG	64	
Db	78	TGGCTGTGTTTGTGTGAAGAAAGATATGCCCATCAAGATGTACATGGAGGGCAACCGTTAATG	137	
QY	65	GCCCACTACTTCAAGTGCAGAGGGCCGAGGGCGAAGGCAACCCCTCGCCCGGCAACCCGAGACA	124	
Db	138	GCCATTATTTTCAAGTGTAGAGAGAGGAGAGACGGGACCCCATTTTACAAGTATGCGAGACA	197	
QY	125	TGAGAAATCCAGTGTGACCGAGGGGCGGCCCCCTTCGCTTCCTGAGCTTTCGACATCTCGGCCCT	184	
Db	198	TGAGGATTCATGTCAACCGAAGGGGCTCCATTATACATTGGCTTCGACATTTTGGGCAACGCT	257	
QY	185	GCTGCGAGTACCGCGAGCAGAGACCTTCGTGACACACACCGCGAGATGCCCGACCTTCTTCA	244	
Db	258	GTTTGAAGTACGGCAGCAGAGACCTTTGTCTCACCATATACGACAGATATCCCGATTCTTCA	317	
QY	245	AGCAGAGCTTCCCCGAGGGCTTACCTGTGGAGAGAAACCAACCTTACGAGAGCGCGGCA	304	
Db	318	AGCAGTCTTTCCTCTGAAAGCTTTTACTTGGGAAAGAACCAACCTATGAAAGATGGAGCA	377	
QY	305	TCCGAGACCGCCCAACCGAGACACACAGCTGTGAGGGGCAACGCTGTATCTTACAAGGTGAAG	364	
Db	378	TTCTTACTCTCATCAGACACACAGCTGTGAGGGGACCTGCTTATATATACAAAGTGAAG	437	
QY	365	TGCAACGCGACCAACTTCCCGCGCGACCGCCCGTGATGAGAAACAAGACGGCGCTGGG	424	
Db	438	TCTTTGTATCCAAATTTTCTGTGTATGGCCCGGTATGAAAGAACAAATCAGAGAGGATGGG	497	
QY	425	AGCCAGCACCGAGGTGTATACCCCGAAGAACGCGGTGTGTGCGCGCGGAACTGATGG	484	
Db	498	AGCCATGACCTGAGGGTGTATATCCAGAGAAATGAGTGTCTGTGTGACGTATGTGATGG	557	
QY	485	CCCTGAAAGTGGGGGACCGGGACCTGATCTGGCACCATCAACAGCTACCGGAGACAGA	544	
Db	558	CCCTTAAAGTGGGTATGTGTCTTTGATCTGCACTCTATATCTTCTTACAGGTCCAAAGA	617	
QY	545	AGGCGGTGCGGCGCCCTGACCAATGCCCCGCTTCCACTTACACGACATCCGCGCTCAGATGC	604	
Db	618	AAGCAGTCCGTGCTTGTGACATATCCAGAGATTTCAATTTTACAGACATCCGCTTCAAGATGC	677	
QY	605	TGCGAGAGAGAGAGCAGATCTTTCAGAGCTGTATCGAGGCGCACGCTGCGCCCGGTACACGC	664	
Db	678	CGAGGAAAAAGAAAGACGATCTTTGAACTGTACGAGACATCTGTGCTAGGTACAGTG	737	
QY	665	ACCTGCCCCGAGAGGCGCAACTGA	687	
Db	738	ATCTTCTCGAAAAAAGCAAAATTGA	760	
RESULT 13				
ABA00806				
ID	ABA00806	standard; cDNA; 760 BP.		
XX	ABA00806;			
XX	AC			
XX	01-APR-2003	(first entry)		
DT	XX			
DE	XX	Multiple mutant Chromoprotein #2 cDNA.		
KW	XX	Gene; kindling fluorescent protein; kindling stimulus; movement;		
KW	XX	labeling; fluorescence resonance energy transfer; FRET;		
KW	XX	bioluminescence resonance energy transfer; BRER; biosensor;		
KW	XX	automated screening; ss.		
OS	XX	Heteractis crispa.		
PH	XX			
FT	XX	Key	Location/Qualifiers	
FT	XX	CDS	77..760	
FT	XX		/*tag= a	
FT	XX		/product= "Multiple mutant Chromoprotein #2"	

Query	Subject	Score	Length	Ident	Mismatches	Indels	Gaps
1	muscle (182, A)	63.6%	760	77.5%	0	0	0
2	muscle (503, T)	63.6%	760	77.5%	0	0	0
3	muscle (549, .550, AT)	63.6%	760	77.5%	0	0	0
4	muscle (594, T)	63.6%	760	77.5%	0	0	0
5	muscle (662, .663, AT)	63.6%	760	77.5%	0	0	0
6	muscle (678, C)	63.6%	760	77.5%	0	0	0
7	muscle (686, A)	63.6%	760	77.5%	0	0	0
8	muscle (686, A)	63.6%	760	77.5%	0	0	0
9	muscle (686, A)	63.6%	760	77.5%	0	0	0
10	muscle (686, A)	63.6%	760	77.5%	0	0	0
11	muscle (686, A)	63.6%	760	77.5%	0	0	0
12	muscle (686, A)	63.6%	760	77.5%	0	0	0
13	muscle (686, A)	63.6%	760	77.5%	0	0	0
14	muscle (686, A)	63.6%	760	77.5%	0	0	0
15	muscle (686, A)	63.6%	760	77.5%	0	0	0
16	muscle (686, A)	63.6%	760	77.5%	0	0	0
17	muscle (686, A)	63.6%	760	77.5%	0	0	0
18	muscle (686, A)	63.6%	760	77.5%	0	0	0
19	muscle (686, A)	63.6%	760	77.5%	0	0	0
20	muscle (686, A)	63.6%	760	77.5%	0	0	0
21	muscle (686, A)	63.6%	760	77.5%	0	0	0
22	muscle (686, A)	63.6%	760	77.5%	0	0	0
23	muscle (686, A)	63.6%	760	77.5%	0	0	0
24	muscle (686, A)	63.6%	760	77.5%	0	0	0
25	muscle (686, A)	63.6%	760	77.5%	0	0	0
26	muscle (686, A)	63.6%	760	77.5%	0	0	0
27	muscle (686, A)	63.6%	760	77.5%	0	0	0
28	muscle (686, A)	63.6%	760	77.5%	0	0	0
29	muscle (686, A)	63.6%	760	77.5%	0	0	0
30	muscle (686, A)	63.6%	760	77.5%	0	0	0
31	muscle (686, A)	63.6%	760	77.5%	0	0	0
32	muscle (686, A)	63.6%	760	77.5%	0	0	0
33	muscle (686, A)	63.6%	760	77.5%	0	0	0
34	muscle (686, A)	63.6%	760	77.5%	0	0	0
35	muscle (686, A)	63.6%	760	77.5%	0	0	0
36	muscle (686, A)	63.6%	760	77.5%	0	0	0
37	muscle (686, A)	63.6%	760	77.5%	0	0	0
38	muscle (686, A)	63.6%	760	77.5%	0	0	0
39	muscle (686, A)	63.6%	760	77.5%	0	0	0
40	muscle (686, A)	63.6%	760	77.5%	0	0	0
41	muscle (686, A)	63.6%	760	77.5%	0	0	0
42	muscle (686, A)	63.6%	760	77.5%	0	0	0
43	muscle (686, A)	63.6%	760	77.5%	0	0	0
44	muscle (686, A)	63.6%	760	77.5%	0	0	0
45	muscle (686, A)	63.6%	760	77.5%	0	0	0
46	muscle (686, A)	63.6%	760	77.5%	0	0	0
47	muscle (686, A)	63.6%	760	77.5%	0	0	0
48	muscle (686, A)	63.6%	760	77.5%	0	0	0
49	muscle (686, A)	63.6%	760	77.5%	0	0	0
50	muscle (686, A)	63.6%	760	77.5%	0	0	0
51	muscle (686, A)	63.6%	760	77.5%	0	0	


```
DB 203 TGAGGATTCATGTCACCGAAGGCGCTCCATTACCATTTGGCTTCGACATTTTGGCACCGT 262
QY 185 GCTGCGAGTACGGCAGAGGACCTTGCTGACACACCGCGAGATCCCGGACTTCTTCA 244
DB 263 GTTTGTAATGACGACAGACCTTTGTCCACATACGCGAGATTTCCGATTTCTTCA 322
QY 245 AGCAGAGCTTCCCGGAGGCTTCACTGTGAGAGAAACAACACTTACGAGACGCGGCA 304
DB 323 AGCAGCTTTTCCCTGAAGCTTACTTGGGAAAGAACACCAACTTGAAGATGGAGGCA 382
QY 305 TCCTGACCGCCACCGACACCAAGCTGAGGGCACTGCTGATCTTACAGGTAAAG 364
DB 383 TTCTTACTGCTCATCAGACACAAGCCTGAGGGGAACTGCTTATATACAAAGTAAAG 442
QY 365 TGCAGCGCACCACTTCCCGCGGAGGCGCCGCGTGAAGAAACAAGAGCGGCGTGGG 424
DB 443 TCCTTGGTACCAATTTTCTGCTGATGAGCCCGTGAAGAAACAATCAGAGGATGG 502
QY 425 AGCCGACGACCGAGTGTGTGTACCCCGAGAACGGCGTGTGCGGCGCGGAAGTATGG 484
DB 503 AGCCATGCACTGAGTGTGTTATCCAGAGATGTGTCCGTGTGACGTAATGTATGG 562
QY 485 CCTTGAAGTGGGCGACCGGCACTGATCTGCCACCACTACACGACTACCGAGCAAG 544
DB 563 CCTTAAAGTCGGTGAATCGTCTTTGATCTGCATCTCTTACTTCTTACAGGTCCAAAG 622
QY 545 AGGCGCTGCGCGGCTGACCATGCGCGGCTTCACATCCGACATCCGCGCTCAAGATG 604
DB 623 AAGCAGTCCGCTGCTTGAACAATGCAAGATTTCAATTTTACGACATCCGCTTCAAGATG 682
QY 605 TGGCGAAGAAGAAGACGAGTACTTTCAGCTGTACGAGGCGACGCTGGCCGGTACAGCG 664
DB 683 CGAGGAAAAGAAAGACGAGTACTTTGAATCTTACGAGAGCATGTGTGGCTAAGTACGTG 742
QY 665 ACCTGCCCGAGAAAGCCCACTGA 687
DB 743 ATCTTCTGAAAAAGCAATTTGA 765

RESULT 15
ABL41182
ID ABL41182 standard; cDNA; 910 BP.
XX
AC ABL41182;
XX
DT 12-AUG-2002 (first entry)
XX
DE H. crispae chromoprotein wild-type base isoform hccp cDNA.
XX
KM Stichodactylidae; chromoprotein; fluorescent; anthozoan; food; PRET;
KM colouring agent; pharmaceutical; cosmetic; immunoassay; biosensor; gene;
KM fluorescence resonance energy transfer; gene expression; hcfp40; ss.
XX
OS Heteractis crispae.
XX
FH key Location/Qualifiers
FT CDS 82..765
FT /tag= a
FT /product= "chromoprotein"
XX
PN MO200230965-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001MO-US032080.
XX
PR 12-OCT-2000; 2000US-0240018P.
XX
PR 16-JUL-2001; 2001US-0306131P.
XX
PA (CLON-) CLONTECH LAB INC.
XX
PI Lukyanov SA, Fradkov AF, Lukyanov KA, Gurskaya NG;
```

```
XX WPI; 2002-444170/47.
DR P-PSDB; ABB08000.
XX
PT Novel nucleic acid encoding Stichodactylidae chromoprotein and its
PT fluorescent mutant useful as coloring agent; labels in analyte detection
PT assays; markers in recombinant DNA applications and filters in
PT sunscreens.
PS
XX
XX Example; Fig 17; 81bp; English.
XX
CC The invention relates to a nucleic acid present in other than its natural
CC environment and encoding an Stichodactylidae chromoprotein or its
CC fluorescent mutant, where the fluorescent protein has an emission maximum
CC ranging from 580-660 nm. The polynucleotides and encoded proteins are
CC useful in applications employing a chromo or fluorescent nucleic acid or
CC protein. Recombinant vectors comprising the nucleic acid is useful for
CC producing an Anthozoan chromo and/or fluorescent protein. The
CC chromoproteins, and their fluorescent mutants are useful as colouring
CC agents capable of imparting colour or pigment to a particular composition
CC of matter. The chromoproteins can be incorporated into a variety of
CC different compositions including food compositions, pharmaceuticals,
CC cosmetics, living organisms, e.g. animals and plants, and as labels in
CC analyte detection assays, e.g. assays for biological analyses of interest
CC (see ABL41167 for a detailed description of the various uses of the
CC chromoproteins). The present sequence represents the cDNA encoding an
CC alternative embodiment of the H. crispae wild-type base isoform hccp cDNA
XX
SQ Sequence 910 BP; 250 A; 199 C; 212 G; 249 T; 0 U; 0 Other;
Query Match 63.3%; Score 435; DB 6; Length 910;
Best Local Similarity 77.3%; Pred. No. 1,6e-67;
Matches 528; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 5 TGAGCGGCTCTGTGAAGAGATATGCGATTAAGATGACTATGAGGGCACTGTGAAG 64
DB 83 TGCTGTGTTTGTGAAGAAAGTATGCGCATMAAGTGTACATGGAAGGACAGGTTAATG 142
QY 65 GCCACTACTTCAAGTGTGCGAGGCGAGGCGGCAAGCCCTTGCAGGCAAGGAGCA 124
DB 143 GCCATTATTTCAAGTGTGAAGAGAGGAGAGAGGCAAGCCCATTTTACAGGTACGAGCA 202
QY 125 TGAGAAATCCACGTGACCGAGGCGGCCCTTCGCTTGCCTTGAATCTGCGCCCT 184
DB 203 TGAGGATTCATGTCACCGAAGGCGGCTCCATTTACATTTGCTTGCATTTTGGCACCGT 262
QY 185 GCTGCGAGTACGGCAGCAGGACCTTCTGTGACACCAACCGCGAGATCCCGACTTTTCA 244
DB 263 GTTGTGAGTACGCGCAGCAGGACCTTTGTCCACCATACGCAAGATTCGCCGATTTCTTCA 322
QY 245 AGCAGAGCTTCCCGGAGGCTTCACTGTGAGAGAAACAACCACTTACGAGAGCGGCGCA 304
DB 323 AGCAGCTTTTCCCTGAAGCTTACTTGTGAAAGAACCAACCACTTGAAGATGGAGGCA 382
QY 305 TCCTGACCGCCACCGACACCAAGCTGAGGGCACTGCTGATCTTACAAAGTGAAG 364
DB 383 TTCTTACTGCTCATCAGACACAAGCCTGAGGGGAACTGCTTATATACAAAGTAAAG 442
QY 365 TGCAGCGCACCACTTCCCGCGGAGGCGCCGCTGATGAAGAAACAAGCGCGCTGGG 424
DB 443 TCCTTGGTACCAATTTTCTGCTGATGAGCCCGTGTGAAGAAACAATCAGAGAGATGG 502
QY 425 AGCCGACGACCGAGGCTGTGTACCCCGAGAACGGCGTGTGCGGCGGGAAGTATGG 484
DB 503 AGCCATGCACTGAGTGTGTTTATCCAGAGAAATGTGTCTGTGTGACGTAATGTATGG 562
QY 485 CCTTGAAGTGGGCGACCGGCACTGATCTGCCACCACTTACCAAGTACCGGAGCAAG 544
DB 563 CCTTAAAGTCGGTGAATCGTCTTTGATCTGCATCTCTTACTTACAGGTCCAAAG 622
QY 545 AGGCGGTGCGGCGCTGACATATGCCGCTTCACTTACCAAGCAATCCGGCTTCAGATGC 604
DB 623 AAGCAGTCCGTGCTTGAACAATGCAAGATTTCAATTTTACAGACATCCGCTTCAGATGC 682
```

Qy	605	TGCGAAGAGAGACGAGTACTTGAAGCTGTACGAGGCCAGCGTGGCCCGGTACAGCG	664
Db	683	CGAGGAAACGAAAGACGAGTACTTGAAGCTGTACGAAACATCTGTGGCTAGGTACAGTG	742
Qy	665	ACCTGCCCCGAGAGGCCCACTGA	687
Db	743	ATCTTCTGAAAAAGCAATTGA	765

Search completed: July 15, 2004, 12:12:28
Job time : 399 secs